(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 July 2003 (17.07.2003)

PCT

(10) International Publication Number
WO 03/058238 A1

(51) International Patent Classification[7]: G01N 33/50, 33/68, C12Q 1/68

(21) International Application Number: PCT/EP03/00196

(22) International Filing Date: 10 January 2003 (10.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02000615.1 10 January 2002 (10.01.2002) EP

(71) Applicant (*for all designated States except US*): MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN [DE/DE]; Berlin (DE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): WECKWERTH, Wolfram [DE/DE]; Kuckuckstrasse 36, 14089 Berln (DE). FIEHN, Oliver [DE/DE]; Hufelandstrasse 20, 10407 Berlin (DE).

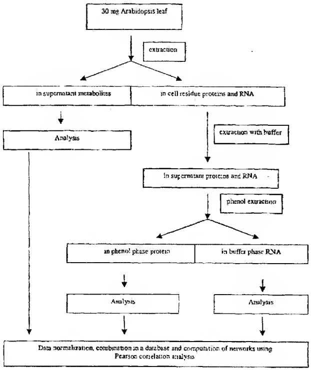
(74) Agent: VOSSIUS & PARTNER; Siebertsrasse 4, 81675 München (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI,

[Continued on next page]

(54) Title: COMBINED METABOLOMIC, PROTEOMIC AND TRANSCRPTOMIC ANALYSIS FROM ONE, SINGLE SAMPLE AND SUITABLE STATISTICAL EVALUATION OF DATA



(57) Abstract: Disclosed is a method for providing data useful for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material, whereby said analysis involves suitable statistical evaluation and correlation analysis on the data obtained, preferably also including network analysis, said method being characterized by the step of extracting, identifying and quantifying at least two compound classes of the group consisting of metabolites, proteins and RNA from at least one sample from said biological source material, wherein the compounds of said at least two classes are each determined from one sample. Furthermore disclosed is a corresponding method for qualitatively analyzing metabolites, proteins and/or RNA in a biological source material comprising providing data on metabolites, proteins and/or RNA, performing suitable statistical evaluation and correlation analysis on the data obtained; and optionally further performing a network analysis. Also disclosed is the use of a mixture of solvents suitable for extracting RNA, proteins and metabolites from a biological sample for extracting metabolites, and optionally also proteins and/or RNA from a biological sample in order to perform metabolite profiling.

WO 03/058238 A1

SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

COMBINED METABOLOMIC, PROTEOMIC AND TRANSCRIPTOMIC ANALYSIS FROM ONE, SINGLE SAMPLE AND SUITABLE STATISTICAL EVALUATION OF DATA

The present invention relates to a method for providing data useful for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material, whereby said analysis involves suitable statistical evaluation and correlation analysis on the data obtained, preferably also including network analysis, said method being characterized by the step of extracting, identifying and quantifying at least two compound classes of the group consisting of metabolites, proteins and RNA from at least one sample from said biological source material, wherein the compounds of said at least two classes are each determined from one sample. The invention furthermore relates to a corresponding method for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material comprising providing data on metabolites, proteins and/or RNA, performing suitable statistical evaluation and correlation analysis on the data obtained; and optionally further performing a network analysis. Moreover, the invention relates to the use of a mixture of solvents suitable for extracting RNA, proteins and metabolites from a biological sample for extracting metabolites, and optionally also proteins and/or RNA from a biological sample in order to perform metabolite profiling.

In the post-genomic era, a functional assignment of genes is getting more and important that goes beyond simple homology search and rough estimations of biochemical and biological roles. Instead, the scientific community tends towards a more comprehensive understanding of biology. Accordingly, the function of a gene is now regarded as depending on developmental and environmental changes as well as on the actual expression level of other genes rather than a simple, linear cause-event relationship. This change in paradigm will necessarily lead to the need to experimentally describe the state of biological tissues in depth on different levels, i.e. not only on the level of transcripts or protein expression, but also on the metabolite level and the interrelatedness of these levels. For example, Fiehn (Comp. Funct. Genom. 2 (2001), 155-168) considers a combination of results from in depth characterization of genetically altered organisms

using transcriptomics, proteomics and metabolomics in order to more comprehensively understand living organisms more feasible than was imagined before. Metabolites are the end products of cellular regulatory processes, and their levels can be regarded as the ultimate response of biological systems to genetic or environmental changes. In parallel to the terms "transcriptome" and "proteome", the set of metabolites synthesized by a biological system compose its "metabolome".

In the growing fields of transcriptomics, proteomics and metabolomics, there is a rapidly ongoing progress in improving the techniques necessary for collecting data and for processing the data obtained. Current strategies and limitations for the quantitative analysis of cellular responses at all of the three levels mRNA, proteins, and metabolites have been recently summarized in a short review by Fiehn (Curr. Opin. Biotechnol. 12 (2001), 82-86) including thoughts on database requirements and informatic tools.

Today, transcriptomic approaches seem to give the best coverage of genome level responses. However, due to limitations in analytical precision and high costs, few transcriptomic studies adequately meet rigid statistical requirements. There are several prominent analytical approaches used for transcript profiling: microarray-based approaches (cDNAs or oligonucleotides) (e.g. reviewed in Granjeaud, Bioessays 21 (1999), 781-790); sequencing-based approaches, such as serial analysis of gene expression (SAGE) (Velculescu, Science 270 (1995), 484-487) or massively parallel signature sequencing (MPSS) (Brenner, Nat. Biotechnol. 18 (2000), 630-634); and differential-display-based approaches, such as arbitrarily primed (AP) PCR (Welsh, Nucl. Acids Res. 20 (1992), 4965-4970) and cDNA-amplified fragment length polymorphism (AFLP) (Bachem, Plant J. 9 (1996), 745-753). Of all of these, microarray technology has received the most widespread interest. This technique has been used extensively in yeast and human work, and thousands of expression profiles have been examined (e.g. Hughes, Cell 102 (2000), 109-126).

For proteomic approaches, two-dimensional gel electrophoresis (2DE) is well established in many biological laboratories (Thiellement, Electrophoresis 20 (1999), 2013-2026) and is comparatively inexpensive. However, if the full set of proteins separated by 2DE gel is to be identified, highly automated systems are needed for cutting spots, digesting proteins, and analyzing peptides using mass spectrometry. Capillary isoelectric focusing is an alternative to 2DE that can be directly coupled to ion cyclotron resonance mass spectrometers (Fourier-transform mass spectrometry, FT-MS) for the analysis of both

crude protein mixtures (Jensen, Electrophoresis 21 (2000), 1372-1380) and complex peptide digests (Gao, J. Microcolumn 12 (2000), 383-390). Further advances have been made with approaches where liquid chromatographic (LC) separation has been combined with mass spectrometrical detection (MS). For instance, LC/LC/MS techniques allow more sensitive and rapid separations of complex peptide mixtures when cation exchange is coupled to reversed phase LC columns. This strategy has been shown to be superior in speed, robustness and protein coverage, when compared with 2DE analysis (Link, Nat. Biotechnol. 17 (1999), 676-682). According to a rather new technique, quantification of protein abundances can be performed using isotope coded affinity tags (ICAT) with precisions as accurate as 12% relative standard deviations (Gygi, Nat. Biotechnol. 17 (1999), 1112-1118). But to date, this technique has not been utilized for proteomic studies that go beyond one-to-one comparative experiments.

Compared to transcriptomic and proteomic approaches, analytical techniques for metabolite detection and quantification are far more robust and mature. Analytical precisions may be below 1% relative standard deviations, and dynamic ranges may exceed four orders of magnitude. For instance in plant research, mass spectrometry has been used for decades to determine metabolic target compounds, but only recently has the idea been pursued of expanding the list of targets in order to profile a limited number of primary metabolites. Moreover, estimates of the total number of metabolites different members of the plant kingdom synthesise, that is estimated of plant metabolome sizes, range from 90,000-200,000. To determine comprehensive metabolomes, therefore, one must cope with the sheer complexity of mass spectra found in chromatograms and apply new technologies for efficient *de novo* compound identification. Most recently, high-througput profiling of metabolic snapshots has been demonstrated for the first time in the context of plant functional genomics (Fiehn, Nat. Biotechnol. 18 (2000), 1157-1161). Profiles of 326 metabolites of two Arabidopsis ecotypes were compared with two single gene mutants. Cluster analysis revealed distinct "metabolic phenotypes" for each of the four genotypes.

On the other hand, evaluation of the profiling data from transcriptome, proteome and metabolome by bioinformatics is also currently in progress. On the basis of the assumption that large sets of genes and proteins follow synchronized patterns, attempts at taming profiling data have been made using clustering algorithms that group raw data in an unbiased way (Eisen, Proc. Natl. Acad. Sci. USA 95 (1998), 14863-14868).

Clustering allows genes, proteins or metabolites to be grouped according to their profiles. All clustering methods used so far, however, have significant drawbacks (Bittner, Nat. Genet. 22 (1999), 213-215) that make them unsuitable for detecting complex relationships in data networks. The inability to detect non-linear correlations is one such limitation. A new way to cover similarities and correlation of expression profiles based on mutual informational entropy has been proposed by Butte (Proc. Natl. Acad. Sci. USA 97 (2000), 12182-12186). Until recently, all clustering methods also lacked a means by which to integrate statistics. Each gene or experiment was mapped into n-dimensional expression space without covering the fact that the sharp point within this space is merely a probability cloud. Hughes (Cell 102 (2000), 109-126) made an attempt to include statistics in their clustering methods so as to see which clusters were trustworthy. In addition, profiling bioinformatics is moving towards methods that try to incorporate as much available knowledge as possible. Recently, two groups used statistical methods to correlate expression profiles with potential promoter sequences (Jensen, Bioinformatics 16 (2000), 326-333; Tavazoie, Nat. Genet. 22 (1999), 281-285). Marcotte (Nature 402 (1999), 83-86) tried to elucidate the functions of unknown ORFs on the basis of five different data sources, including expression profiling data. Brown (Proc. Natl. Acad. Sci USA 97 (2000), 262-267) used a special neuronal network to identify potential gene functions on the basis of the expression profile and the transcriptional patterns of well-annotated genes.

Yet, with regard to the goal of comprehensively understanding living organisms, there is still an ongoing need for improved data mining tools, and better tools for integrating the results of experimentally determined molecular phenotypes with predictions made by computational simulations of cellular networks (Fiehn, Comp. Funct. Genom. 2 (2001), 155-168).

In view of the above outlines, it is evident that the developments in obtaining comprehensive data on the levels of RNA, proteins and metabolites and the attempts to derive therefrom new insights into the complex regulation of gene expression appear to be quite promising, however, also require further improvements.

Thus, the technical problem underlying the present invention is to provide a method that allows it to improve the meaningfulness of correlation data on gene expression and metabolite states of a given biological source material.

This technical problem is solved by the provisions of the embodiments as characterized in the claims.

Accordingly, the present invention relates to a method for providing data useful for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material, whereby said analysis involves suitable statistical evaluation and correlation analysis on the data obtained, preferably also including network analysis,
said method being characterized by the step of extracting, identifying and quantifying at least two compound classes of the group consisting of metabolites, proteins and RNA from at least one sample from said biological source material, wherein the compounds of said at least two classes are each determined from one sample.

The particular achievement of the present invention lies in the fact that at least two of the three compound classes RNA, proteins and metabolites are each extracted and analyzed from one single sample. Accordingly, the phrase "the compounds of said at least two classes are each determined from one sample" means that, for each sample collected in the course of applying the method of the invention, the compounds of all the compound classes analyzed are extracted from one and the same sample.
The method of the present invention may be carried out altogether taking one sample or taking more than one, preferably a multitude of samples. For example, to provide data from one sample may be applicable if corresponding reference data is already available from other samples, e.g. in a database. Data from a series of samples can for instance be useful for investigating a process along a time course, such as a developmental process. On the other hand, samples taken from different genotypes of organisms being at the same developmental stage may likewise deliver useful data, e.g., for characterizing the influence of a particular genotype on gene expression, protein composition and/or metabolite composition.

The method of the present invention represents a clear improvement over prior art techniques for providing data, e.g. for network analyses, since there data on RNA levels, protein levels and/or metabolite states were each sampled individually, i.e. determined in different samples. Thus, aiming at combining data sets from e.g. two compound classes hitherto was hampered by the uncertainty of whether the individual data sets actually represent corresponding cellular states. Since cellular states may vary depending on developmental or environmental conditions or on the genotype, it is often uncertain whether such data sets to be combined are indeed compatible, i.e. so that any correlations measured, e.g. between metabolite data and gene expression data, indeed reflect the actual processes in the cells under investigation. In addition, prior art attempts to combine data from different compound classes may suffer from calibration problems. In particular, these occur when one wants to directly set absolute values from one data set into relation to absolute values from another data set. To do so would require to have control values within each data set that allows it to calibrate the data. These drawbacks, i.e. the uncertainty of potential non-compatibility and calibration problems, are largely overcome or at least substantially reduced by the provisions of the present invention. By deriving the quantitative data on two or three different compound classes each from one single sample, it is guaranteed that the data sets obtained reflect just the same cellular status of the biological source material. Also calibration should present a decisively reduced problem if the data sets are obtained according to the method of the invention. Since the amount of material extracted is the same for each compound class analyzed of one sample and the data on the individual compound classes are derived from the identical cells, it can be expected that absolute values of different compound classes are directly correlatable without need of calibration. Still if, due to a certain way of subsequently extracting the different compounds from one sample, the absolute values finally obtained will not be directly correlatable, the person skilled in the art knows how to relate the values obtained to the amount of starting material extracted so that the absolute values may resume their direct correlatability.

The method of the present invention relates to the provision of quantitative data on at least two compound classes in a biological source material. The term "compound class" relates to a group of compounds which is either metabolites, proteins or RNA. In a preferred embodiment, data on all three compound classes are obtained by said method.

Also preferred is the combination of metabolites with either proteins or RNA. The term "metabolite" refers to any substance within a certain biological source material that is non-peptidic and not a nucleic acid molecule. Preferably, the metabolites addressed by the present invention rather have a lower molecular weight, i.e. for instance not more than 4000 Da, preferably not more than 2000 Da, more preferably not more than 1000 Da. Typically, the metabolites to be analyzed belong to the following, however non-limiting list of compounds: carbohydrates (e.g. sugars, oligo- and polysaccharides such as polyglucans as for example starch or polyfructans), sugar alcohols, amines, amino alcohols, aliphatics, aliphatic alcohols, amino acids, lipids, fatty acids, fatty alcohols, organic acids, organic phosphates, organic or anorganic ions, nucleotides, sugar nucleotides, sterols, terpenes, terpenoids, flavons and flavonoids, glucosides, carotenes, carotenoids and cofactors.

The term "proteins" refers to any molecules comprising amino acids connected via peptide bonds that are present in the biological source material.

The term "RNA" refers to any ribonucleic acids that can be present in cells contained in the biological source material. Preferred RNA is mRNA, i.e. transcripts from protein-encoding genes. The present method may also be applied by detecting other RNA species than mRNA such as rRNA, tRNA or viral RNA.

The term "biological source material" means any material being or containing living matter, such as cells, tissues, organs or organisms. The method is not restricted to any taxon. Thus, prokaryotes such as archaebacteria or eubacteria (gram-positive or gram-negative) as well as eukaryotes such as yeasts, fungi, plants (i.e. algae or land plants, in particular flowering plants) or animals as for instance insects or vertebrates, especially mammals, and corresponding cell cultures are within its reach. As a preferred biological source material, plant tissue, e.g. leaves, can be used. The feasibility of the method of the invention has first been demonstrated at Arabidopsis leaf (see Example). Other preferred biological source materials may be the classical objects of genetics, such as Drosophila, E.coli, yeast, especially Saccharomyces cerevisiae or C. elegans where lots of mutants are described and regulatory networks integrating, e.g. metabolites and gene expression data, are promising to reveal new insights into gene regulation, e.g. during development. In a further important aspect, the method of the invention may help to improve drug discovery or cancer research. In this context, corresponding biological source material would for example be derivable from animals such as vertebrates, typically birds or

mammals. In connection to this, preferred examples of biological examples may be mammalian, preferably human cell lines or tissue from test animals such as laboratory mice or rats.

The term "sample" encompasses any amount of material taken from the biological source material that is susceptible to the method of the invention. For instance, a sample can be fresh material such as a tissue explant, a body fluid or an aliquot from a bacterial or cell culture, preferably deprived of the culture medium, that may be directly subjected to extraction. On the other hand, samples may also be stored for a certain time period, preferably in a form that prevents destructive action of inherent enzymatic activity, e.g. frozen, for instance in liquid nitrogen, or lyophilized.

Extraction should be carried out so that the compounds one is interested in are dissolved as completely and quantitatively as possible and in a manner that they can later be identified and quantified using appropriate methods. The term "extracting" thereby refers to contacting the material containing the compounds of interest with an extractant (e.g. solvent or mixture of solvents) so that these compounds can be dissolved, followed by separating the solution from the undissolved matter. The person skilled in the art is capable of choosing an appropriate extraction protocol that is suited to isolate the two or three compound classes of interest and which is useful for the biological source material the sample is taken from. In the prior art, methods are described that may be especially useful within the context of the present invention, such as subsequent application of pure solvents or methanol/water mixtures, or other methods such as microwave extraction, ultrasonication, or accelerated solvent extraction.

Usually, it is advantageous to use an extractant that contains or is an organic (i.e. lipophilic) solvent so that the proteins are denatured upon addition of the extractant so that other compounds, i.e. metabolites and/or RNA will not be enzymatically degraded or modified. Stopping of inherent enzymatic activity can also be achieved by freeze clamping, immediate freezing in liquid nitrogen, or by acidic treatments using perchloric or nitric acid (ap Rees and Hill, 1994). Although advantageous for extraction of amines (Bouchereau et al. 2000), acidic treatments may possibly pose problems for many subsequent analytical methods. In a preferred way, frozen samples can be directly extracted by immediately adding organic solvents and applying heat, thereby also inhibiting the recovery of enzymatic activity. Extracting frozen samples that still contain

the original amount of water can be advantageous for applications including analysis of metabolites when compared to extracting lyophilized samples, since lyophilisation may potentially lead to the irreversible adsorption of metabolites on cell walls or membranes. In cases where it is desired to distinguish between metabolite levels in different compartments, samples can be lyophilized prior to non-aqueous fractionation methods (Gerhardt and Heldt 1984, Farré et al. 2001). An alternative approach to non-aqueous fractionation is the use of nuclear magnetic resonance analyses (NMR) to distinguish steady state concentrations of metabolites in different compartments in vivo (Roberts 2000). For tissue culture, a cold shock can be used in which the liquids are infused into cold methanol. All devices needed for further sample preparation should then be kept at cold temperatures (Gonzalez et al. 1997). In addition, polar organic solvents like methanol, methanol-water mixtures, or ethanol can be directly added to freshly frozen tissues (Johansen et al. 1996, Streeter and Strimbu 1998), with an additional step of using non-polar solvents such as chloroform to exhaustively extract lipophilic components. In order to enhance the extraction efficiency, additional energy may be put into the system either directly by heat (e.g. 70°C), or by other techniques such as pressurized liquid extraction (Benthin et al. 1999), supercritical fluid extraction (Jarvis and Morgan 1997, Blanch et al. 1999, Castioni et al. 1995), sonication (Sargenti and Vichnewski 2000), subcritical water extraction (Gàmiz-Gracia and de Castro 2000), microwave techniques (Namiesnik and Gorecki 2000), or pervaporation (Starmans and Nijhuis 1996).

In a preferred embodiment of the method of the invention, extracting comprises the steps of:

(a) extracting the metabolites from said sample with at least one solvent or mixture of solvents;

(b) extracting the proteins from the remainder of the sample after step (a);

(c) extracting the RNA from the remainder of the sample after step (a); and

(d) optionally dissolving remaining cellular material contained in said sample.

This sequence of steps describes the isolation of all three compound classes. If, however, one is interested in data of only two compound classes, these steps may be adapted correspondingly, for instance by omitting step (a) or (b) or (c), depending on whether metabolites, protein or RNA shall not be analyzed. The sequence of steps (a) to

(d) is not necessarily chronological. In particular, step (c) may for instance also be carried out before step (b) or simultaneously.

In step (a), the sample is extracted with an extractant that comprises at least a solvent or a mixture of solvents. The extractant may contain additional substances such as antioxidants, detergents or buffering agents. The term "solvent" refers to the common meaning as used in the prior art, while "aqueous solvent" refers to water or an aqueous buffer.

In a particularly preferred embodiment, step (a) is carried out with a mixture of solvents that comprises at least one highly polar solvent, at least one less polar solvent and at least one lipophilic solvent.

The terms "highly polar", "less polar" and "lipophilic" in connection with solvent has a clear meaning to the person skilled in the art. Usually, the polarity of a solvent is defined by reference to the dielectric constant (ε) of the solvent. This quantity is temperature-dependent, often it is given for solvents at 20°C. In connection with the method of the invention, "highly polar" solvents have a dielectric constant in the range from above 35 to 90, preferably from 36 to 81. "Less polar" solvents have a dielectric constant in the range of above 5.5 to 35, preferably from 6 to 33. "Lipophilic" solvents have a dielectric constant in the range from 1.0 to 5.5, preferably 1.8 to 5.0. Typical examples of highly polar solvents are acetonitrile, N,N-dimethylacetamide, N,N-dimethylformamide, dimethyl sulfoxide, glycerol, nitromethane and water. Typical expamples of less polar solvents are acetic acid, acetone, 2-butanol, 1,2-dichloroethane, dichloromethane, ethanol, ethyl acetate, methanol, 1-propanol, 2-propanol, pyridine and tetrahydrofuran. Furthermore, typical examples of lipophilic solvents are benzene, carbon tetrachloride, chloroform, cyclohexane, diethyl ether, 1,4-dioxane, heptane, hexane, pentane, tetrachloroethylene and toluene.

Preferably, the mixture of highly polar, less polar and lipophilic solvents has one phase. Accordingly, the solvents should be chosen so that, at the particular conditions of extraction applied, the mixture of solvents used has one phase. More preferably, the mixture of solvents for use in step (a) comprises water, methanol and chloroform, most preferably these solvents are contained in the mixture in the approximate proportion by volume of 1:2.5:1

In the method of the invention, the temperature under which extraction is carried out can be crucial to successfully extract a broad range of compounds, in particular metabolites. In the experiments underlying the present invention, a cold temperature has been shown to be favorable in this regard. Accordingly, it is preferred that extraction, in particular the above-mentioned step (a) of the preferred extraction scheme, is carried out at a temperature in the range between -60°C and +4°C, more preferably between -40°C and -10°C, still more preferably between –20°C and –13°C and most preferably at –16°C. Practically, extraction should be carried out at a temperature where the solvents used are liquid.

Steps (a) to (d) of the above-described preferred embodiment of the method of the invention can be carried out according to techniques known to the person skilled in the art and described in the literature. In order to illustrate the sequence of the extraction steps, Figure 1 shows a specific scheme containing steps (a) to (c).

Generally, in step (a), the metabolites are extracted from the sample. For this purpose, the cells or tissue may be broken up using appropriate techniques such as homogenization. After incubating the sample material for sufficient time to dissolve the metabolites quantitatively in the extractant, the solution containing the metabolites should be separated from undissolved material (remainder of the sample).

In steps (b) and (c), the proteins and RNA may be extracted from the remainder of the sample after step (a), using a suitable extractant, e.g. a buffered aqueous solution. The resulting extract can then be subjected to a further separation step, wherein proteins are separated from RNA. This can, for instance, be done by phenol extraction, i.e. by adding phenol, mixing and centrifuging with the result that RNA is in the aqueous phase and proteins in the phenol phase. An overview of various RNA isolation techniques is described in Sambrook and Russell (2001) and Gassen and Schrimpf (1999).

Optionally, as step (d), dissolving remaining cellular material contained in the sample may also be performed in order to provide additional compounds for quantitative analysis. Such undissolved material may comprise membranes or cell wall material that can be dissolved using suitable means such as hydrolyzing enzymes. The hydrolysates may for instance be added to the metabolite fraction or may be analyzed individually.

In a preferred embodiment, the method of the invention further comprises removing detection-disturbing compounds from the metabolites, the polypeptides and/or the RNA prior to identifying and quantifying the metabolites, proteins and/or RNA. This embodiment refers in particular to carbohydrates or other compounds that disturb identification and quantification of RNA. Routinely, compounds that may disturb the detection of the individual compound classes are removed by suitable techniques known to the skilled practitioner if such a removal improves the quality and significance of the detection (i.e. identification and quantification). For example, it has been shown that the presence of carbohydrates disturbs the detection of RNA by microarrays and that the removal of the carbohydrates from the sample may significantly improve the quality of the detected signals.

Identification and quantification of the compounds of interest extracted from the sample can be done according to well-known techniques known in the prior art. For each of the compound classes in question, techniques are described that allow identification and quantification in one step and, moreover, are suited to record the respective compounds contained in the extracts in a comprehensive manner.

For example, metabolites may be identified and quantified using gas chromatography/mass spectrometry (GC/MS), liquid chromatography/mass spectrometry (LC/MS), NMR or FT-IR or combinations thereof. Further useful methods include LC/UV, refractory index determination, the use of radioactivity in connection with suitable methods known to the skilled person, thin layer chromatography (TLC), capillary electrophoresis (CE), CE/UV, CE/laser induced fluorescence (LIF), fluorescence detection, electrochemical detection (i.e. colorimetry), direct injection MS, flow injection MS, MS/MS, MS/MS/MS, and further combinations of MS steps (MS^n), fourier transform ion mass spectrometry (FT/MS), and gel permeation chromatography (GPC).

An exemplary non-biased analysis is described in Fiehn et al. (2000). With the aim of functionally characterizing plant mutants, detection and relative quantifications of 326 distinct compounds (ranging from primary polar metabolites to sterols) was carried out for both identified and non-identified compounds, after normalization to internal references and plant tissue fresh weights. Different plant mutants were compared to the corresponding parental genotypic backgrounds, and the data were used for statistical analysis as well as for defining metabolic phenotypes that were derived from clustering

tools. Another example of GC/MS analyses that can be applied in the method of the invention has been described by Roessner et al. (2001), who used it for comprehensively studying the metabolism in potato tubers. Alternatively, metabolite data can be obtained by extended chromatographic analysis as described by Tweeddale et al. (1998) where, after growing wild type and mutant E.coli strains in minimal media and ^{14}C-labelled glucose, 70 metabolites could be separated using two dimensional thin layer chromatography. The relative quantification of metabolites was carried out by radioactive detection.

Suitable techniques for identifying and quantifying proteins are known to the person skilled in the art and described in the literature. They include for example LC/MS and two-dimensional electrophoresis and protein staining. A comprehensive overview of different strategies for protein extraction and proteome analysis from various tissue that may be useful in connection with the method of the invention can be found in "2D-Proteome Analysis Protocol" by Andrew Link (1999). For example, the use of 2D-gels combined with mass spectrometry, usually MALDI-TOF, allows the detection and identification of a large number of proteins from a sample in a quality that facilitates the comparison of different protein profiles (Nock et al. 1998; Eggeling et al. 1998). For applications that require a unbiasedness, resolution and/or reproducibility that exceeds the potential of two-dimensional gel electrophoresis (2DE) (e.g. as discussed in Gygi et al. 2000), there are alternative techniques available in the prior art. For instance, Wall et al. have shown that protein mixtures can be fractionated using isoelectric focusing in the liquid phase (Wall et al. 2000, 2001). By coupling to monolithic columns and digesting the protein prior to mass spectrometry, it is possible to achieve the identification of hundreds of proteins. For the purpose of high throughput approaches relative protein quantification as described by Gygi et al. (1999b) may be the technique of choice. When analyzing proteins from yeast, Gygi and co-workers were able to achieve a repeatability better than 12% relative standard deviation. This method involves linking protein Cys-residues to stable isotopically labeled chemicals that include biotin moieties for sample purification and preconcentration. This method, also referred to as isotope-coded affinity tag (ICAT), has been successfully applied in the experiments made in connection with the present invention (see Example and Figures 4B and C). Protein identification and quantification may likewise be done by de novo peptide sequencing (Goodlett et al. 2001). In addition, when applying LC/MS-based methods it may be beneficial to increase the run times of

liquid chromatography in order to ensure that only a few peptides per time interval reach the mass spectrometer (Washburn et al. 2001). This may be achieved by stepwise elution of the peptide mixture from strong cation exchangers onto nanoscale reverse phase columns prior to ion trap MS/MS experiments. By applying this strategy, it was possible to reach a genome coverage of ~25 percent for yeast. In total, 1500 proteins were detected, ranging from low abundant transcription factors to proteins with up to 14 transmembrane domains, and high and low abundant enzymes. Most importantly, it could be proven that no bias against codon usage was found in the detected proteins. With regard to MS and subsequent MS/MS experiments, it may be recommendable to use mass spectrometers with ultimate sensitivity and mass resolution. Here, for example, a combination of nanoLC columns (Li et al. 2001, Shen et al. 2001a, 2001b) or capillary isoelectric focusing (Jensen et al. 2000) prior to fourier-transform ion cyclotron mass spectrometry (FT-MS) can be of use. By applying either of these techniques, up to 10^6 peptides per run can theoretically be separated, with up to 50,000 peptides found experimentally. Another strategy feasible for the method of the invention may be labeling approaches using stable isotopes. Such a method was originally developed by Aebersold and then adapted to proteomic experiments (Smith et al. 2001, Goshe et al. 2001). Using such a method, it was possible to identify and quantify up to a thousand proteins in 4 h runs. Moreover, by using infrared laser photodissociation, peptides can be fragmented within the cyclotron in order to confirm peptide identification that is for example solely based on accurate masses (Conrads et al. 2000).

RNA can be identified and quantified using a number of techniques currently available in the field such as hybridization on nylonfilters, cDNA microarrays, DNA chips loaded with oligonucleotides (see Granjeaud, Bioessays 21 (1999), 781-790), serial analysis of gene expression (SAGE; Velculescu, Science 270 (1995), 484-487), massively parallel signature sequencing (MPSS; Brenner, Nat. Biotechnol. 18(6) (2000), 630-634) differential-display-based approaches, such as arbitrarily primed (AP) PCR (Welsh, Nucl. Acids Res. 20 (1992), 4965-4970) and cDNA-amplified fragment length polymorphism (AFLP) (Bachem, Plant J. 9 (1996), 745-753). Systematic approaches for a comprehensive mRNA expression profiling that may be applied in connection with the method of the invention are described in Hughes et al. (2000) and Lockhart and Winzeler (2000).

The term "quantitatively analyzing metabolites, proteins and/or RNA" refers to any mathematical analysis method that is suited to further process the quantitative data provided by the method of the invention. This data represents the amount of the compounds analyzed present in each sample either in absolute terms (e.g. weight or moles per weight sample) or in relative terms (i.e. normalized to a certain reference quantity).

Quantitative analysis involves suitable statistical evaluation and correlation analyses. The former includes normalization to the total content of the respective compounds within one compound class, correction of background levels and the combination of the data sets obtained from different compound classes into a single data sheet. Corresponding mathematical methods and computer programs are known to the practicioner. Examples include SAS, SPSS and systatR. As the next step, the statistically pre-treated data may be subjected to a pairwise correlation analysis. Here series of pairs of data points from the analyzed compounds are looked at for correlation, whether positive or negative, for instance using Pearson's correlation coefficient. In a preferred embodiment, the quantitative analysis referred to in the method of the invention furthermore involves network analysis. Network analysis aims at finding out higher order interplays of multiple factors on the basis of pairwise correlation data. By taking several data sets each obtained from one sample, correlations between metabolites, proteins, and/or RNA as well as among these classes of compounds can be analyzed in order to derive information about the network regulation of biological systems, e.g. upon genetic or environmental perturbation. A comprehensive overview of methods for quantitatively analyzing data obtained according to the method of the invention including principle component analysis, "snapshot analysis", pearson correlation analysis, mutual information and network analyses can be found in Fiehn (2001).

The present invention furthermore relates to a method for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material comprising

(a) providing data on metabolites, proteins and/or RNA in said biological source material according to the method described above;

(b) performing suitable statistical evaluation and correlation analysis on the data obtained; and

(c) optionally further performing a network analysis on the data obtained in step (b).

The features of this method have already been described above in connection with the method for providing data and likewise apply in the context of this aspect of the invention.

In addition, the present invention relates to the use of a mixture of solvents comprising at least a highly polar solvent, at least a less polar solvent and at least a lipophilic solvent as described herein above for extracting metabolites from a sample of a biological source material in order to perform metabolite profiling.
It has been shown that such a mixture of solvents as it is characterized in its various aspects above is better suited for extracting metabolites than other extraction methods described in the prior art such as subsequent application of pure solvents or methanol/water mixtures, or other methods such as microwave extraction, ultrasonication, or accelerated solvent extraction. Thus, it is possible to benefit from the special extraction strength of such a mixture of solvents for metabolite profiling. The term "metabolite profiling" refers to quantitatively analyzing metabolites according to the explanations given above.

According to a preferred embodiment of this use, additionally proteins and/or RNA is extracted from said sample.

These and other embodiments are disclosed and encompassed by the description and examples of the present invention. Further literature concerning any one of the methods, uses and compounds to be employed in accordance with the present invention may be retrieved from public libraries, using for example electronic devices. For example the public database "Medline" may be utilized which is available on the Internet, for example under http://www.ncbi.nlm.nih.gov/PubMed/medline.html. Further databases and addresses, such as http://www.ncbi.nlm.nih.gov/, http://www.infobiogen.fr/, http://www.fmi.ch/biology/research_tools.html, http://www.tigr.org/, are known to the person skilled in the art and can also be obtained using, e.g., http://www.google.de. An overview of patent information in biotechnology and a survey of relevant sources of patent information useful for retrospective searching and for current awareness is given in Berks, TIBTECH 12 (1994), 352-364.

Furthermore, the term "and/or" when occurring herein includes the meaning of "and", "or" and "all or any other combination of the elements connected by said term".

The present invention is further described by reference to the following non-limiting figures and examples.

<u>The Figures show:</u>

Figure 1: Scheme illustrating the steps performed in the experiment of the Example.

Figure 2: A: GC/MS (TOF) direct analysis of hydrophilic and lipophilic metabolites.
B: LC/MS/MS analysis of the metabolites.

Figure 3: A: SDS PAGE, lanes 1 to 3 show the proteins from three different samples extracted according to the scheme shown in Figure 1.
B: Agarose gel electrophoresis, lines 1 to 9 show RNA from nine different samples extracted according to the scheme shown in Figure 1.

Figure 4: A: Above: LC/MS/MS analysis of peptides obtained upon trypsination of protein extracted from Arabidopsis leaf according to the method described. A list of the proteins identified in the LC/MS/MS analysis is given in Table 3.
B and C: ICAT analysis of proteins extracted according to the scheme shown in Figure 1 and explained in the Example from an Arabidopsis sample and an Arabidopsis control sample showing the relative quantification of the proteins (for details see Gygi et al. 1999b). B: list of peptides in one fraction after cation-exchange- and avidin-affinity chromatography. C: MALDI TOF spectrum with enlarged detail of a peptide pair for relative quantification

Figure 5: Metabolite network of Pearson correlations using 30 samples.

Figure 6: Analysis of metabolites, proteins, and transcripts extracted from a single Arabidopsis leaf sample.

A Functional characterisation of identified proteins from a single Arabidopsis Col2 leaf sample. Majority of the proteins are chloroplast-related.

B Comparison of the integrative extraction protocol with a conventional plant RNA extraction kit

C RNA blot analysis of Arabidopsis isopropylmalate synthase (IPMS) transcripts (two isoforms, GI9758 and GI9759360) in three replicate samples of Arabidopsis Col2 leafs extracted according to the scheme in Figure 1.

Figure 7: Principle component and hierarchical cluster analysis of metabolites and proteins.

A Integrated principle component analysis (PCA) of metabolites and proteins reveals complete clustering of the two genotypes

B Individual loadings of the metabolites and proteins for two different Arabidopsis varieties C24 and Col2.

C Hierarchical cluster analysis (HCA) of a merged metabolite-protein dataset. (for details see text). Abbreviations of proteins and metabolites: RUBISCO activase: ribulose-1,5-bisphosphate carboxylase/oxygenase activase (S04048); RUBISCO: ribulose-1,5-bisphosphate carboxylase/oxygenase (NP_051067); GAPDH: glyceraldehyde-3-phosphate dehydrogenase (AAD10209); asc peroxidase: L-ascorbate peroxidase (S20866); put kinase: protein kinase, putative (At3g24550); SUC: sucrose; FUC: fucose; SHI: shikimate; CP12 like: CP12 protein precursor-like protein (At3g62410); ATPsynthase: ATP synthase CF1 beta chain (NP_051066); peroxidase: peroxidase, putative (At3g49120); put protein: protein, putative (At3g63190); put TK: transketolase - like protein (At3g60750), put aldolase: putative fructose-bisphosphate aldolase (AF428455_1); put oxidase: glycolate oxidase (At3g14420); GST protein: spindly (gibberellin signal transduction

protein) (At3g11540); ATPase: ATPase alpha subunit (NP_051044); EF-1: translation elongation factor eEF-1 alpha chain (gene A4) (S08534); catalase: catalase (AAB07026); put protein: protein, putative (At3g47140); P-protein like: (At4g33010); put protein: protein, putative (At3g57190); PS I RC: putative photosystem I reaction center subunit II precursor (At1g03130); PS I SUI: photosystem I subunit III precursor (CAB52747); CIT: citrate; XYL: xylose; TRE: trehalose; SIN: sinapinate; GAL: galacturonate; CITMA: citramalate; ASP: aspartate; GLUC: gluconate; MAL: malate; INO: inositol; FUM: fumarate.

The following Examples illustrate the invention:

Experimental set-up

Plant material

Arabidopsis thaliana plants were cultivated in phytotrons under highly controlled light, gas, and temperature conditions assuring approximately identical environmental conditions for each plant sample. Biological variation among independent samples of the same genotypes is attributed to the inherent fluctuation of the biochemical network due to slightly changed environments.

Extraction procedure

30-100 mg Arabidopsis leaf sample at a developmental stage of 1.1 according to Boyes et al. (2001) were harvested and immediately frozen in liquid nitrogen. Tissue was homogenized under liquid nitrogen using a retsch mill. A one-phase solvent mixture of methanol/chloroform/water 2.5:1:1 (v/v/v) was kept at –20 °C and 2 mL was added to the tissue and thoroughly mixed at 4 °C for 30 min to precipitate proteins and DNA/RNA and to disassociate metabolites from membrane and cell wall components. After centrifugation, the remaining pellet consisting of DNA/RNA, proteins, starch, membranes, and cell wall components was extracted in a second step with 1 mL methanol/chloroform

1:1 (v/v) at –20°C. The organic solvent extracts were combined and further used for metabolite analysis via GC-TOF. For that purpose the chloroform phase was separated from the water/methanol phase by adding 500 µl water. The resulting water/methanol phase now contained all hydrophilic metabolites such as sugars, amino acids and organic acids, and the chloroform phase all the lipophilic compounds, lipids, chlorophyll, and waxes. The remaining white pellet was further partitioned according to the scheme in Figure 1. The pellet was extracted with 1 mL extraction buffer (0,05 M Tris, pH7,6; 0,5 % SDS; 1 % ß-Mercaptoethanol) and 1 mL water saturated phenol for 1h at 37°C. After centrifugation at 14,000g the remaining pellet was used for cell wall synthesis (data not shown). The phenol phase was separated from the buffer phase and the proteins were precipitated with ice-cold acetone in –20 °C overnight, washed three times with ethanol and dried at room temperature. Remaining protein in the RNA-buffer phase was precipitated with 200 µL chloroform. After centrifugation and separation of the buffer phase, 40µL of acetic acid and 1 mL ethanol were added to precipitate the RNA at 4 °C for 30 min. The pellet was washed with one volume 3 M sodium acetate, and two times with one volume 70% ethanol. The remaining pellet was dissolved in 100 µL RNAse-free water. Amounts and purity of RNA were checked by absorbance at 260 nm and gel electrophoresis in agarose. Construction of Arabidopsis isopropyl-malate synthase (IPMS) probes for hybridisation and Northern blots were performed using standard protocols.

GC-TOF analysis

For GC-TOF analysis, the organic phase was dried and dissolved in 50µl of methoxamine hydrochloride (20mg/mL pyridine) and incubated at 30 °C for 90 min with continuous shaking. Then 80 µL of N-Methyl-N-trimethylsilyltrifluoroacetamid (MSTFA) was added to derivatize polar functional groups at 37°C for 30 min. The derivatized samples were stored at room temperature for 120 min before injection. GC-TOF analysis was performed on a HP 5890 gas chromatograph with standard liners and splitless injection at 230°C injector temperature. The GC was operated at constant flow of 1 mL/min Helium and a 40 m 0.25 mm ID 0.25 µm RTX-5 column with 10 m integrated pre-column. The temperature gradient started at 80°C, was held isocratic for 2 min, and subsequently ramped at 15°C/min to a final temperature of 330°C which was held for 6

min. 20 spectra per second were recorded between m/z 85 to 500. After data acquisition was finished, reference chromatograms were defined that had a maximum of detected peaks over a signal/noise threshold of 20 and used for automated peak identification based on mass spectral comparison to a standard NIST 98. Automated assignments of unique ions for each individual metabolite were taken as default as quantifiers, and manually corrected where necessary. All artifactual peaks caused by column bleeding or phtalates and polysiloxanes derived from MSTFA hydrolysation were removed from the results table. All data were normalized to plant mg FW and to the internal references and log-transformed. t-test, correlation analysis, and variance analysis were performed in Excel 5.

Two-dimensional liquid chromatography/mass spectrometry.

The dried protein pellet was dissolved in freshly prepared 1M Urea in 0.05 M Tris buffer pH 7.6. The complex protein mixture was digested with modified trypsin (Böhringer Mannheim) according to the manufacturer's instructions. The tryptic digest was dried down and dissolved in 300 µl water (1% formic acid). Unsoluble material was removed by centrifugation. An aliquot of the digest (~100 µg protein) was injected onto two-dimensional chromatography on a thermofinnigan proteomeX system coupled to an LCQDecaXp ion trap (Thermofinnigan). The chromatographic separation was done according to manufacturer's instructions. After a 12 cycle run the MS/MS spectra were searched against an Arabidopsis thaliana database (downloaded from the TAIR homepage www.arabidopsis.org) using Turbosequest implemented in Bioworks 3.0 (Thermofinnigan). Matches were filtered according to Wolters et al. (2001). Additionally, we used the multiple scoring filter of Bioworks 3.0 with 50 percent ion coverage. For the quantification approach aliquots of the complex tryptic digest of Arabidopsis leaf protein (50 µL) were analysed on reversed phase chromatography. Quantification was achieved by integrating peak areas of target peptides representative for proteins. These peak areas were normalised to the sum of internal standard peptides that had been added to the mixture (Chelius et al. (2002), Bondarenko et al. (2002)).

Statistical analysis

All quantitative metabolite and protein data were normalized to FW. Principle component analysis (PCA) and hierarchical cluster analysis (HCA) pattern recognition was performed according to Fiehn et al. (2000) using Pirouette software (Infometrix, Woodinville, WA). The integrative data set of metabolites and proteins was log10 transformed. The HCA was performed using Euclidian distances and complete linkage grouping. Variance analyses were performed in MS Excel 5.0.

Example 1: Network analysis of compounds in Arabidopsis leaf

A sample of an *Arabidopsis thaliana* leaf (30mg FW) was extracted with a cold (-16°C) mixture of chloroform / methanol / water (1 : 2.5 : 1 v/v/v). This mixture is a one-phase solution and proved to have an improved extraction strength for metabolites in a comprehensive scale. Identification and relative quantitation of these metabolites was done with GC/MS and LC/MS analysis (see Figure 2A and 2B). In a subsequent step, proteins and RNA were isolated from the remaining cell residue by buffer/phenol extraction and phase separation (see Figure 3A and 3B). The protein was precipitated from the phenol phase with methanol/acetate, washed three times and dried. After tryptic digestion the proteins were identified and relatively quantitated by LC/MS and LC/MS/MS (see Figure 4). RNA was precipitated from the aqueous phase with ethanol and quantitated by gel electrophoresis (see Figure 3B).

Last, all data were normalized to the total content of the respective gene products, corrected for background levels, and combined in a single data sheet. For each individual data set, or for the combined data set, network calculation can be done.

As an example, all pair-wise correlations for a metabolite dataset of 30 samples were analyzed for Pearson's correlation coefficients and visualized in a network (see Figure 5).

Example 2: Correlation profiling of proteins and metabolites of two Arabidopsis genotypes

For each replicate, 30-100 mg FW leaf tissue of an individual Arabidopsis thaliana plant was extracted at 4°C with chloroform/methanol/water (1:2.5:1 v/v/v) according to the

scheme illustrated in Figure 1. This single-phase mixture proved to have improved extraction strength for metabolites in comparison to the former extraction protocol utilizing a methanol/water mixture for 15 min at 70°C. However, when chloroform was left out of the cold extraction mix, i.e. if methanol/water extraction mixtures were used at -20°C, a strong decrease in sucrose content was detected in subsequent GC/TOF analyses, concomitant with a sharp increase in fructose and glucose contents. This indicates that chloroform may inhibit sucrose cleaving enzyme activity, e.g. by invertase or sucrose synthase, by precipitating these enzymes. Total metabolite analysis was performed with GC/TOF (Weckwerth et al., 2001) (see Figure 2A) enabling the detection and quantification of 652 metabolites (see Table 1). Replication of metabolite analysis revealed a high recovery and a mean coefficient of variance (CV) of 10%. In a subsequent step, proteins and mRNA were isolated from the remaining cell residue using buffer/phenol extraction and phase separation (see Figure 1). In Figure 6B, a comparison of this method with a conventional RNA extraction method is shown. The extraction procedure achieves a higher level of RNA recovery than does a typical RNA isolation kit extraction (see Experimental set-up, supra) with 30% CV in 28 samples. To test the utility of the mRNA for hybridisation we analysed the expression of isopropylmalate synthase (IPMS) from Arabidopsis (see Figure 6C). The average amount of total protein extracted according to the scheme in Figure 1 was 1.3 mg per 100mg FW with 17% CV. The overall extraction process resulted in good recovery of metabolites, proteins, and transcripts. After complete extraction, the remaining cell pellet was used for cell wall analysis giving rise to clear and typical cell wall profiling (data not shown).

The protein fraction was analysed using shotgun proteomics (Wolters et al., 2001; Washburn et al., 2001; Koller et al., 2002). The complex mixture of the tryptic Arabidopsis leaf protein digest was analysed via two-dimensional capillary liquid chromatography and tandem mass spectrometry on an ion trap mass spectrometer (LCQ Deca Xp Plus) and a subsequent database search performed using TurboSequest implemented in ThermoFinnigan Bioworks 3.0. In a single Arabidopsis Col2 leaf sample extracted according to the scheme in Figure 1 586 peptides and 297 corresponding proteins have been identified using very stringent criteria to avoid false positives (see Experimental set-up (supra) and Table 2). A classification of the detected proteins from one sample is shown in Figure 6A. We applied the integrative extraction process to two Arabidopsis genotypes, C24 and Col2, to test if we are able to determine different biochemical

phenotypes and general biochemical patterns using this strategy. C24 and Col2 showed an overlap of 153 proteins (see Table 2). The data-dependent detection of peptides was strongly contingent on the estimated abundance of the corresponding proteins in the digest such as RUBISCO (Pang et al., 2002). Thus, the high number of non-overlapping proteins also indicates differences in the protein-profiles of these different Arabidopsis genotypes. A set of 22 proteins appearing in both varieties was chosen for the quantification approach. These proteins were quantified by integrating their corresponding peptide areas in a one-dimensional LC/MS analysis. These areas were normalized to internal standard peptides as described in Chelius et al. (2002) and Bondarenko et al. (2002). The analytical precision was tested by adding internal standard peptides to the sample. The deviation of the internal standards – in other words the technical variation of the extraction process, stability of electrospray and matrix effects – was ~25 % CV. Each genotype was represented by ten independent samples. The relative integrals of the peptides in each sample were normalised to the fresh weight (FW) of the corresponding sample.

The metabolites in the corresponding samples were identified and quantified with GC/TOF. A list of all identified metabolites is given in Table 1. Fourteen of the most abundant metabolites were normalized to the FW and combined with the data of the quantified proteins to form an integrative dataset. Most important for the integrative analysis of such heterogeneous data is the appropriate transformation. A homogeneous dataset was achieved by applying log10 transformation (see statistical analysis, supra).

The first step of analysis is to test if we are able to discriminate the two genotypes suggesting different biochemical phenotypes under the same environment. We applied principle component analysis (PCA) according to Fiehn et al. (2000). Both Col2 and C24 were completely separated into genotype-clusters (Figure 7A and 7B). In contrast to NMR or other fingerprinting methods, the individual identification of compounds by our method enables the investigation of distinct metabolite-protein cross-correlations in a multitude of samples. The aim is to reveal hierarchical structures within complex biochemical networks depending on the genotype-phenotype relationship (Cooper et al., 2002). Based on the detection of these fundamental correlations in an integrative dataset – for instance a distance or a Pearsons Matrix – it is possible to expose instantaneous causal connectivities in a regulatory network representing a snapshot of the actual state of the system (Arkin et al., 1997; Kell et al., 2000; Vance et al., 2002). To make use of

such a refined analysis it is important to differentiate biological variability and technical measurement error. The quantified proteins showed an overall variability of ~39% whereas individual variation was up to 70% exceeding clearly the overall analytical precision of ~25% CV. The same was observed for the metabolites according to Fiehn et al. (2000). We calculated the ratio of standard deviation to the mean for every variable, metabolite and protein. These ratios appeared not to be correlated to the means ($r_{metabolites}$= 0.38 and $r_{proteins}$=0.23), indicating that the relative variation of these compounds does not depend on their abundance. This is indicative of high biological variation among independent samples, even samples collected from tissues at seemingly identical developmental stage and grown under highly controlled environmental conditions. In Figure 7C, a hierarchical analysis of the set of quantified proteins and metabolites is shown. C24 and Col2 datasets are merged together to detect biochemical patterns conserved for both genotypes. A strongly conserved pattern for both varieties is detected for Calvin cycle enzymes such as ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO) and 3-glyceraldehyde dehydrogenase (GAPDH) which is in agreement with the literature (Krapp et al., 1993). Metabolites included in this cluster are sucrose (SUC) and fucose (FUC) suggesting the coordination of sucrose synthesis and degradation and photosynthetic activity. Surprisingly, ascorbate peroxidase is integrated into the Calvin cycle/sucrose cluster giving hints to the connectivity of the oxidative state and carbohydrate metabolism in plants. Inside the metabolite cluster, biochemically related structures such as malate (MAL) and fumarate (FUM) and carbohydrates form subclusters as expected.

Table 1 **summarizing the metabolites of known structure that were identified in the leaf tissue extracts as described in Example 2**

Entry	Name	Sum formula	Mol. weight	Substance class
1	C7 TMS	C10H22O2Si	202	fatty acid time standard
2	C9 TMS	C12H26O2Si	230	fatty acid time standard
4	C11 TMS	C14H30O2Si	258	fatty acid time standard
5	C13 TMS	C16H34O2Si	286	fatty acid time standard
6	C15 TMS	C18H38O2Si	314	fatty acid time standard
7	C19 TMS	C22H46O2Si	370	fatty acid time standard
8	C23 TMS	C26H54O2Si	426	fatty acid time standard
9	C27 TMS	C30H62O2Si	482	fatty acid time standard
10	C31 TMS	C34H70O2Si	538	fatty acid time standard
11	RIBITOL TMS			sugar alcohol quantify standard
239	1HO-14:0 TMS	C17H38OSi	286	alcohol
240	1HO-16:0 TMS	C19H42OSi	314	alcohol
241	1HO-18:0 TMS	C21H46OSi	342	alcohol
242	1HO-20:0 TMS	C23H50OSi	370	alcohol
243	1HO-22:0 TMS	C25H54OSi	398	alcohol
267	1HO-26:0 TMS	C29H62OSi	454	alcohol
274	1HO-26:0 TMS	C29H62OSi	454	alcohol
268	1HO-28:0 TMS	C31H66OSi	482	alcohol
269	1HO-29:0 TMS	C32H68OSi	496	alcohol
270	1HO-30:0 TMS	C33H70OSi	510	alcohol
254	1HO-31:0 TMS	C34H72OSi	524	alcohol
271	1HO-32:0 TMS	C35H74OSi	538	alcohol
256	1HO-33:0 TMS	C36H76OSi	552	alcohol
308	ETHYLENEGLYCOL,O,O-TMS	C8H22O2Si2	206	alcohol
167	GLYCEROL 3TMS	C12H32O3Si3	308	alcohol
324	PROPAN-1,2-DIOL,O,O-TMS	C9H24O2Si2	220	alcohol
293	OXAMIDE,N,N-TMS	C8H20N2O2Si2	232	amid
296	OXAMIDE,N,N-TMS MEOX	C9H23N3O2Si2	261	amid
130	1,3-DIAMINOPROPANE,N,N,N,N-TMS	C15H42N2Si4	362	amine
303	CADAVERINE,N,N,N,N-TMS	C17H46N2Si4	390	amine
306	DIETHANOLAMINE,N,O,O-TMS	C13H35NO2Si3	321	amine
307	DIETHANOLAMINE,O,O-TMS	C10H27NO2Si2	249	amine
331	ETHANOLAMINE,N,N,O-TMS	C11H31NOSi3	277	amine
321	PHENETHYLAMINE,N-TMS	C14H27NSi2	265	amine
310	PUTRESCINE,N,N,N,N-TMS	C16H44N2Si4	376	amine
309	SPERMIDINE,N,N,N,N,N-TMS	C22H59N3Si5	505	amine
186	2-AMINO-ADIPINIC ACID 3TMS	C15H35NO4Si3	377	amino acid
322	2-METHYLSERINE,N,O,O-TMS	C13H33NO3Si3	335	amino acid
320	2-METHYLSERINE,O,O-TMS	C10H25NO3Si2	263	amino acid
15	ALANINE,N,O-TMS	C9H23NO2Si2	233	amino acid
36	ALLOTHREONINE,N,O,O-TMS	C13H33NO3Si3	335	amino acid
22	ASPARAGINE,N,N,N,O-TMS	C16H40N2O3Si4	420	amino acid
21	ASPARAGINE,N,N,O-TMS	C13H32N2O3Si3	348	amino acid
126	ASPARTIC ACID,N,O,O-TMS	C13H31NO4Si3	349	amino acid
145	B-ALANINE TMS	C12H31NO2Si3	305	amino acid
304	CITRULLINE,O,N,N,N-TMS	C15H37N3O3Si3	391	amino acid
127	CYSTATHIONINE,N,N,O,O-TMS	C19H46N2O4SSi4	510	amino acid
132	CYSTEINE,N,O,S-TMS	C12H31NO2SSi3	337	amino acid
29	CYSTIN TMS			amino acid
168	GLUTAMIC ACID 3TMS	C14H33NO4Si3	363	amino acid
28	GLUTAMINE,N,N,N,O-TMS	C17H42N2O3Si4	434	amino acid
27	GLUTAMINE,N,N,O-TMS	C14H34N2O3Si3	362	amino acid
24	GLYCINE,N,N,O-TMS	C11H29NO2Si3	291	amino acid
265	GLYCINE,N,O-TMS	C8H21NO2Si2	219	amino acid
30	HOMOCYSTEINE,N,N,O-TMS	C13H33NO2SSi3	351	amino acid
159	HOMOGLUTAMINE TMS			amino acid
35	HOMOSERINE 3TMS	C13H33NO3Si3	335	amino acid
13	ISOLEUCINE,N,O-TMS	C12H29NO2Si2	275	amino acid
153	L-ARGININMONOHYDROCHLORID TMS			amino acid
12	LEUCINE, N,O-TMS	C12H29NO2Si2	275	amino acid
33	L-HYDROXYPROLINE,N,O,O-TMS	C14H33NO3Si3	347	amino acid

32	LYSINE,N,N,N',O-TMS	C18H46N2O2Si4	434	amino acid
14	METHIONINE,N,O-TMS	C11H27NO2SSi2	293	amino acid
312	NORLEUCINE,N,N,O-TMS	C15H37NO2Si3	347	amino acid
316	NORLEUCINE,N,O-TMS	C12H29NO2Si2	275	amino acid
317	NORLEUCINE,O-TMS	C9H21NO2Si	203	amino acid
313	NORVALINE,N,N,O-TMS	C14H35NO2Si3	333	amino acid
318	NORVALINE,N,O-TMS	C11H27NO2Si2	261	amino acid
319	NORVALINE,O-TMS	C8H19NO2Si	189	amino acid
16	ORNITHINE,N,N,N',O-TMS	C17H44N2O2Si4	420	amino acid
291	OXAMIC ACID,N,O-TMS MEOX1	C9H22N2O3Si2	262	amino acid
292	OXAMIC ACID,N,O-TMS MEOX2	C9H22N2O3Si2	262	amino acid
332	OXAMIC ACID,O-TMS MEOX	C6H14N2O3Si	190	amino acid
34	PHENYLALANINE,N,O-TMS	C15H27NO2Si2	309	amino acid
314	PIPECOLIC ACID,N,O-TMS	C12H27NO2Si2	273	amino acid
315	PIPECOLIC ACID,O-TMS	C9H19NO2Si	201	amino acid
17	PROLINE,N,O-TMS	C11H25NO2Si2	259	amino acid
169	PYROGLUTAMIC ACID 2TMS	C11H23NO3Si2	273	amino acid
197	SACCHAROPINE TMS			amino acid
19	SERINE,N,O,O-TMS	C12H31NO3Si3	321	amino acid
20	SERINE,O,O-TMS	C9H23NO3Si2	249	amino acid
334	S-METHYLCYSTEINE	C16H41NO2SSi4	423	amino acid
336	S-METHYLCYSTEINE,N,N,O-TMS	C13H33NO2SSi3	351	amino acid
323	S-METHYLCYSTEINE,N,O-TMS	C10H25NO2SSi2	279	amino acid
31	THREONINE,N,O,O-TMS	C13H33NO3Si3	335	amino acid
25	TRYPTOPHANE,N,N',O-TMS	C20H36N2O2Si3	420	amino acid
26	TRYPTOPHANE,N,O-TMS	C17H28N2O2Si2	348	amino acid
18	VALINE,N,O-TMS	C11H27NO2Si2	261	amino acid
311	PUTRESCINE + CO2 4TMS	C17H44N2O2Si4	420	artef. of putrescine
286	DOPAMINE,N,N,O,O-TMS	C20H43NO2Si4	441	catecholamine
284	NORADRENALINE,N,N,O,O,O-TMS	C23H51NO3Si5	529	catecholamine
285	NORMETHYLADRENALINE 4TMS	C21H45NO3Si4	471	catecholamine
333	TYRAMINE,N,O-TMS	C14H27NOSi2	281	catecholamine
74	CELLOBIOSE MEOX1 TMS			disaccharide
75	CELLOBIOSE MEOX2 TMS			disaccharide
69	ISOMALTOSE MEOX1 TMS			disaccharide
70	ISOMALTOSE MEOX2 TMS			disaccharide
189	LACTOSE MEOX1 TMS			disaccharide
190	LACTOSE MEOX2 TMS			disaccharide
71	LACTULOSE TMS			disaccharide
134	LAMINARIBIOSE MEOX1 TMS			disaccharide
135	LAMINARIBIOSE MEOX2 TMS			disaccharide
65	MALTOSE MEOX1 TMS			disaccharide
66	MALTOSE MEOX2 TMS			disaccharide
72	MELIBIOSE MEOX1 TMS			disaccharide
73	MELIBIOSE MEOX2 TMS			disaccharide
149	NIGEROSE MEOX1 TMS			disaccharide
150	NIGEROSE MEOX2 TMS			disaccharide
76	PALATINOSE TMS			disaccharide
79	RAFFINOSE TMS			disaccharide
68	SUCROSE TMS			disaccharide
67	TREHALOSE TMS			disaccharide
77	TURANOSE MEOX1 TMS			disaccharide
78	TURANOSE MEOX2 TMS			disaccharide
194	XYLOBIOSE MEOX1 TMS			disaccharide
195	XYLOBIOSE MEOX2 TMS			disaccharide
221	3 HO-10:0 ME TMS	C14H30O3Si	274	fatty acid
220	3 HO-10:1 ME TMS	C14H28O3Si	272	fatty acid
223	3 HO-12:0 ME TMS	C16H34O3Si	302	fatty acid
222	3 HO-12:2 ME TMS	C16H30O3Si	298	fatty acid
224	3 HO-14:2 ME TMS	C18H34O3Si	326	fatty acid
230	3 HO-14:3 ME TMS	C18H34O3Si	326	fatty acid
225	3 HO-14:4 ME TMS	C18H32O3Si	324	fatty acid
217	3 HO-6:0 ME TMS	C10H22O3Si	218	fatty acid
218	3 HO-8:0 ME TMS	C12H26O3Si	246	fatty acid
219	3 HO-8:1 ME TMS	C12H24O3Si	244	fatty acid
182	CIS-10-HEPTADECENOIC ACID ME	C18H34O2	282	fatty acid
178	CIS-10-PENTADECENOIC ACID ME	C16H30O2	254	fatty acid
209	CIS-11,14,17-EICOSATRIENOIC ACID ME	C21H36O2	320	fatty acid
206	CIS-11-EICOSENOIC ACIDME	C21H40O2	324	fatty acid
213	CIS-13,16-DOCOSADIENOIC ACID ME	C23H42O2	350	fatty acid

212	CIS-13-DOCOSENOIC ACID ME	C23H44O2	352	fatty acid
215	CIS-15-TETRACOSENOIC ACID ME	C25H48O2	380	fatty acid
202	CIS-6,9,12-OCTADECATRIENOIC ACID ME	C19H32O2	292	fatty acid
208	CIS-8,11,14-EICOSATRIENOIC ACID ME	C21H36O2	320	fatty acid
204	CIS-9,12,15-OCTADECATIENOIC ACID ME	C19H32O2	292	fatty acid
201	CIS-9,12-OCTADECADIENOIC ACID ME	C19H34O2	294	fatty acid
180	CIS-9-HEXADECENOIC ACID ME	C17H32O2	268	fatty acid
183	CIS-9-OCTADECENOIC ACID ME	C19H36O2	296	fatty acid
176	CIS-9-TETRADECENOIC ACID ME	C15H28O2	240	fatty acid
211	DOCASANOIC ACID ME	C23H46O2	354	fatty acid
172	DODECANOIC ACID ME	C13H26O2	214	fatty acid
205	EICOSANOIC ACID ME	C21H42O2	326	fatty acid
210	HENEICOSANOIC ACID ME	C22H44O2	340	fatty acid
181	HEPTADECANOIC ACID ME	C18H36O2	284	fatty acid
179	HEXADECANOIC ACID ME	C17H34O2	270	fatty acid
170	NONADECANOIC ACID ME	C20H40O2	312	fatty acid
200	OCTADECANOIC ACID ME	C19H38O2	298	fatty acid
171	PENTACOSANOIC ACID ME	C26H52O2	396	fatty acid
177	PENTADECANOIC ACID ME	C16H32O2	256	fatty acid
216	TETRACOSANOIC ACID ME	C25H50O2	382	fatty acid
175	TETRADECANOIC ACID ME	C15H30O2	242	fatty acid
203	TRANS-9,12-OCTADECADIENOIC ACID ME	C19H34O2	294	fatty acid
184	TRANS-9-OCTADECENOIC ACID ME	C19H36O2	296	fatty acid
214	TRICOSANOIC ACID ME	C24H48O2	368	fatty acid
174	TRIDECANOIC ACID ME	C14H28O2	228	fatty acid
264	UNDECANOIC ACID ME	C12H24O2	200	fatty acid
266	CIS-11,14-EICOSADIENOIC ACID ME	C21H38O2	322	fatty acid
273	3HO-16:1 ME TMS	C20H40O3Si	356	hydroxy-carboxylic acid
259	3HO-16:2 ME TMS	C20H38O3Si	354	hydroxy-carboxylic acid
260	3HO-16:3 ME TMS	C20H36O3Si	352	hydroxy-carboxylic acid
272	3HO-16:0 ME TMS	C20H42O3Si	358	hydroxy-carboxylic acid
125	PHOSPHORIC ACID,O,O,O-TMS	C9H27O4PSi3	314	inorganic acid
278	D-MANNONO-DELTA-LACTAM 4TMS	C18H43NO5Si4	465	lactam
277	MANNONO-DELTA-LACTAM MEOX 4TMS	C19H46N2O5Si4	494	lactam
62	ARABINOSE MEOX1 4TMS	C18H45NO5Si4	467	monosaccharide
63	ARABINOSE MEOX2 4TMS	C18H45NO5Si4	467	monosaccharide
64	ARABINOSE MEOX2 4TMS	C18H45NO5Si4	467	monosaccharide
50	ERYTHROSE MEOX1 3TMS	C14H35NO4Si3	365	monosaccharide
51	ERYTHROSE MEOX2 TMS	C14H35NO4Si3	365	monosaccharide
46	FRUCTOSE MEOX1 5TMS	C22H55NO6Si5	569	monosaccharide
47	FRUCTOSE MEOX2 5TMS	C22H55NO6Si5	569	monosaccharide
58	FUCOSE MEOX1 4TMS	C19H47NO5Si4	481	monosaccharide
59	FUCOSE MEOX2 4TMS	C19H47NO5Si4	481	monosaccharide
56	GALACTOSE MEOX1 TMS	C22H55NO6Si5	569	monosaccharide
57	GALACTOSE MEOX2 TMS	C22H55NO6Si5	569	monosaccharide
141	GLUCOHEPTULOSE MEOX1 TMS			monosaccharide
142	GLUCOHEPTULOSE MEOX2 TMS			monosaccharide
138	GLUCOSAMINE MEOX1 TMS			monosaccharide
137	GLUCOSAMINE MEOX2 TMS			monosaccharide
41	GLUCOSE MEOX1 5TMS	C22H55NO6Si5	569	monosaccharide
42	GLUCOSE MEOX2 5TMS	C22H55NO6Si5	569	monosaccharide
37	GLYCERALDEHYD MEOX1 2TMS	C10H25NO3Si2	263	monosaccharide
38	GLYCERALDEHYD MEOX2 TMS	C10H25NO3Si2	263	monosaccharide
279	LYXOSE,O,O,O,O-TMS MEOX1	C18H45NO5Si4	467	monosaccharide
280	LYXOSE,O,O,O,O-TMS MEOX2	C18H45NO5Si4	467	monosaccharide
45	MANNOSE MEOX TMS	C22H55NO6Si5	569	monosaccharide
60	RHAMNOSE MEOX1 4TMS	C19H47NO5Si4	481	monosaccharide
61	RHAMNOSE MEOX2 4TMS	C19H47NO5Si4	481	monosaccharide
52	RIBOSE MEOX1 4TMS	C18H45NO5Si4	467	monosaccharide
53	RIBOSE MEOX2 4TMS	C18H45NO5Si4	467	monosaccharide
54	XYLOSE MEOX1 4TMS	C18H45NO5Si4	467	monosaccharide
55	XYLOSE MEOX2 4TMS	C18H45NO5Si4	467	monosaccharide
192	XYLULOSE MEOX 4TMS	C18H45NO5Si4	467	monosaccharide
107	2-OXOBUTYRIC ACID MEOX1 TMS	C8H17NO3Si	203	organic acid
160	2-OXOVALERIC ACID MEOX2 TMS	C9H19NO3Si	217	organic acid
120	3-HYDROXYPYRUVIC ACID MEOX1 TMS	C10H23NO4Si2	277	organic acid
121	3-HYDROXYPYRUVIC ACID MEOX2 TMS	C10H23NO4Si2	277	organic acid
122	3-OXOGLUTARIC ACID MEOX1 TMS	C12H25NO5Si2	319	organic acid
123	3-OXOGLUTARIC ACID MEOX2 TMS	C12H25NO5Si2	319	organic acid
166	4-AMINOBUTYRIC ACID 3TMS	C13H33NO2Si3	319	organic acid

119	4-HYDROXYBENZOIC ACID,O,O-TMS	C13H22O3Si2	282	organic acid
288	6-HYDROXYNICOTINIC ACID,O,O-TMS	C12H21NO3Si2	283	organic acid
148	ACONITIC ACID 3TMS	C15H30O6Si3	390	organic acid
109	ALPHA-KETOGLUTARIC ACID MEOX1 2TMS	C12H25NO5Si2	319	organic acid
110	ALPHA-KETOGLUTARIC ACID MEOX2 2TMS	C12H25NO5Si2	319	organic acid
99	ASCORBIC ACID TMS			organic acid
101	BENCOIC ACID TMS	C10H14O2Si	194	organic acid
97	CITRIC ACID TMS	C18H40O7Si4	480	organic acid
282	FERULIC ACID,O,O-TMS	C16H26O4Si2	338	organic acid
95	FUMARIC ACID TMS	C10H20O4Si2	260	organic acid
111	GALACTURONIC ACID MEOX1 5TMS	C22H53NO7Si5	583	organic acid
112	GALACTURONIC ACID MEOX2 5TMS	C22H53NO7Si5	583	organic acid
115	GLUCARIC ACID 1,4LACTONE MEOX1 TMS			organic acid
116	GLUCARIC ACID 1,4LACTONE MEOX2 TMS			organic acid
147	GLUCOHEPTONIC ACID TMS			organic acid
198	GLUCOHEPTONIC ACID TMS			organic acid
114	GLUCONIC ACID LACTONE TMS			organic acid
113	GLUCONIC ACID,O,O,O,O,O,O-TMS			organic acid
187	GLUCURONIC ACID MEOX1 5TMS	C22H53NO7Si5	583	organic acid
188	GLUCURONIC ACID MEOX2 5TMS	C22H53NO7Si5	583	organic acid
163	GLUCURONIC ACID-E-LACTONE MEOX TMS			organic acid
117	GLUTARIC ACID,O,O-TMS	C11H24O4Si2	276	organic acid
118	GLYCERIC ACID,O,O,O-TMS	C12H30O4Si3	322	organic acid
330	GLYCOLIC ACID,O,O-TMS	C8H20O3Si2	220	organic acid
162	GLYOXYLIC ACID MEOX TMS	C6H13NO3Si	175	organic acid
262	GLYOXYLIC ACID MEOX TMS	C6H13NO3Si	175	organic acid
263	GLYOXYLIC ACID MEOX TMS	C6H13NO3Si	175	organic acid
329	HEXANOIC ACID,O-TMS	C9H20O2Si	188	organic acid
124	HYDROXYBUTANOIC ACID,O,O-TMS	C10H24O3Si2	248	organic acid
165	INDOLE-3-ACETIC ACID 2TMS	C16H25NO2Si2	319	organic acid
139	ISOCITRIC ACID TMS	C18H40O7Si4	480	organic acid
328	LACTIC ACID,O,O-TMS	C9H22O3Si2	234	organic acid
152	LAURIC ACID TMS	C15H32O2Si	272	organic acid
93	MALIC ACID TMS	C13H30O5Si3	350	organic acid
261	NICOTINIC ACID TMS	C9H13NO2Si	195	organic acid
98	OXALIC ACID TMS	C8H18O4Si2	234	organic acid
105	PYRUVIC ACID MEOX TMS	C7H15NO3Si	189	organic acid
104	QUINIC ACID TMS	C22H52O6Si5	552	organic acid
143	SACCHARIC ACID TMS			organic acid
106	SALICYLIC ACID 2TMS	C13H22O3Si2	282	organic acid
96	SHIKIMIC ACID TMS	C19H42O5Si4	462	organic acid
281	SINAPINIC ACID,O,O-TMS	C17H28O5Si2	368	organic acid
133	SUCCINIC ACID 2TMS	C10H22O4Si2	262	organic acid
100	TARTARIC ACID,O,O,O,O-TMS	C16H38O6Si4	438	organic acid
289	THREONIC ACID,O,O,O,O-TMS	C16H40O5Si4	424	organic acid
156	2-METHYLMALIC ACID 3TMS	C14H32O5Si3	364	organic acid
227	2-OXOISOCAPROIC ACID MEOX1 TMS	C10H21NO3Si	231	organic acid
157	2-OXOISOCAPROIC ACID MEOX2 TMS	C10H21NO3Si	231	organic acid
158	2-OXOISOVALERIC ACID MEOX1 TMS	C9H19NO3Si	217	organic acid
228	2-OXOISOVALERIC ACID MEOX2 TMS	C9H19NO3Si	217	organic acid
229	2-OXOVALERIC ACID MEOX1 TMS	C9H19NO3Si	217	organic acid
131	3-HYDROXYPYRIDINE,O-TMS	C8H13NOSi	167	organic compound
283	4-HYDROXYPYRIDINE,O-TMS	C8H13NOSi	167	organic compound
299	ALLANTOIN,N,N,N,N,N-TMS	C19H46N4O3Si5	518	organic compound
298	ALLANTOIN,N,N,N,N-TMS	C16H38N4O3Si4	446	organic compound
300	ALLANTOIN,N,N,N-TMS	C13H30N4O3Si3	374	organic compound
297	URIC ACID,N,N,O,O-TMS	C17H36N4O3Si4	456	organic compound
151	2-PHOSPHOGLYCERATE TMS	C15H39O7PSi4	474	phosphorylated compound
39	3-PHOSPHOGLYCERATE TMS	C15H39O7PSi4	474	phosphorylated compound
49	FRUCTOSE-6-PHOSPHATE MEOX2 TMS			phosphorylated compound
48	FRUCTOSE-6-PPHOSPHATE MEOX1 TMS			phosphorylated compound
43	GLUCOSE-6-PHOSPHATE MEOX1 TMS			phosphorylated compound
44	GLUCOSE-6-PHOSPHATE MEOX2 TMS			phosphorylated compound
128	GLUCURONIC ACID-6-PHOSPHATE TMS			phosphorylated

				compound
40	GLYCERO-3-PHOSPHATE 4TMS	C15H41O6PSi4	460	phosphorylated compound
305	GLYCEROL-2-PHOSPHATE,O,O,O,O-TMS	C15H41O6PSi4	460	phosphorylated compound
144	RIBOSE-5-PHOSPHATE TMS			phosphorylated compound
196	SORBITOL-6-PHOSPHATE TMS			phosphorylated compound
146	SUCROSE-6-PHOSPHATE TMS			phosphorylated compound
155	CYTOSINE 2TMS			pyrimidine
164	METHYLCYTOSINE 2TMS	C11H23N3OSi2	269	pyrimidine
154	THYMINE 2TMS	C11H22N2O2Si2	270	pyrimidine
235	ALPHA-TOCOPHEROL 1TMS	C32H58O2Si	502	steroid
237	BETA-SITOSTEROL 1TMS	C32H58OSi	486	steroid
234	CAMPESTEROL 1TMS	C31H56OSi	472	steroid
232	CHOLESTEROL 1TMS	C30H54OSi	458	steroid
233	LANOSTEROL 1TMS	C33H58OSi	498	steroid
236	STIGMASTEROL 1TMS	C32H56OSi	484	steroid
88	ERYTHRITOL TMS	C16H42O4Si4	410	sugar alcohol
90	GALACTITOL TMS			sugar alcohol
92	INOSITOL,O,O,O,O,O,O-TMS	C24H60O6Si6	612	sugar alcohol
185	LACTITOL TMS			sugar alcohol
89	MALTITOL TMS			sugar alcohol
91	MALTOTRITOL TMS			sugar alcohol
86	MANNITOL TMS			sugar alcohol
193	ONONITOL TMS	C22H54O6Si5	554	sugar alcohol
87	SORBTOL TMS			sugar alcohol
140	XYLITOL 5TMS	C20H52O5Si5	512	sugar alcohol
84	CELLOTRIOSE MEOX1 TMS			trisaccharide
85	CELLOTRIOSE MEOX2 TMS			trisaccharide
83	ISOMALTOTRIOSE TMS			trisaccharide
80	MALTOTRIOSE MEOX1 TMS			trisaccharide
81	MALTOTRIOSE MEOX2 TMS			trisaccharide
82	MELEZITOSE TMS			trisaccharide
275	1,4-DIDEOXY-1,4-IMINO-D-ARABINITOL 3TMS	C14H35NO3Si3	349	
276	1,4-DIDEOXY-1,4-IMINO-D-ARABINITOL 3TMS	C14H35NO3Si3	349	
325	2-AMINO-2-METHYL-1,3-PROPANDIOL 3TMS	C13H35NO2Si3	321	
302	3-METHYLAMINOL-1,2-PROPANDIOL,N,O,O-TMS	C13H35NO2Si3	321	
301	3-METHYLAMINOL-1,2-PROPANDIOL,O,O-TMS	C10H27NO2Si2	249	
290	THREONINOL,O,O-TMS	C13H35NO2Si3	321	
199	UREA 2TMS	C7H20N2OSi2	204	

Table 2 providing a list of the proteins identified in extracts from C24 and Col2 leaf tissue as described in Example 2

ProteinID	Function	Peptides identified	Found in
NP_051067	ribiose 1,5-bisphosphate carboxylase/oxygenase large chain	23	Col2/C24
NP_051066	ATP synthase CF1 beta chain	10	Col2/C24
S04048	ribulose-bisphosphate carboxylase activase (EC 6.3.4.-) precursor	8	Col2/C24
NP_051044	ATPase alpha subunit	13	Col2/C24
AAL38341	chlorophyll a/b-binding protein	5	Col2/C24
CAB40384	16 kDa polypeptide of oxygen-evolving complex	9	Col2/C24
At3g01500	carbonic anhydrase, chloroplast precursor	8	Col2/C24
At3g60750	transketolase - like protein	6	Col2/C24
At4g10340	light-harvesting chlorophyll a/b binding protein	7	Col2/C24
GCST_MESCR	Aminomethyltransferase, mitochondrial precursor (Glycine cleavage system T protein) (GCVT)	7	Col2/C24
AAD10209	glyceraldehyde 3-phosphate dehydrogenase A subunit	4	Col2/C24
At2g30860	glutathione transferase, putative	4	Col2/C24
At1g06680	photosystem II oxygen-evolving complex 23 (OEC23)	5	Col2/C24
AT5g25980	myrosinase TGG2	5	Col2/C24
At5g35630	glutamate-ammonia ligase (EC 6.3.1.2) precursor, chloroplast	4	Col2/C24
S08534	translation elongation factor eEF-1 alpha chain (gene A4)	3	Col2/C24
At3g55800	sedoheptulose-bisphosphatase precursor	3	Col2/C24
At1g03130	putative photosystem I reaction center subunit II precursor	4	Col2/C24
AAM12979	chlorophyll a/b-binding protein CP29	4	Col2/C24
At3g26060	putative peroxiredoxin	4	Col2/C24
NP_051054	photosystem II protein D2	3	Col2/C24
At2g40840	glycosyl hydrolase family 77 (4-alpha-glucanotransferase)	1	Col2/C24
BAB08951	2-cys peroxiredoxin-like protein	3	Col2/C24
AAM62639	unknown	2	Col2/C24
At1g12900	putative calcium-binding protein, calreticulin	1	Col2/C24
At1g44575	photosystem II 22kDa protein, putative	4	Col2/C24
At4g05180	oxygen-evolving complex protein 16, chloroplast precursor (OEC16)	4	Col2/C24
NP_051072	cytochrome f	3	Col2/C24
AF217459_1	heat shock protein 70	2	Col2/C24
At1g64290	hypothetical protein	1	Col2/C24
At5g26000	glycosyl hydrolase family 1, myrosinase precursor	3	Col2/C24

At3g57190	putative protein	1	Col2/C24
AAK64040	unknown protein	4	Col2/C24
AF428455_1	putative fructose-bisphosphate aldolase	1	Col2/C24
At1g32060	phosphoribulokinase precursor	2	Col2/C24
BAA20945	beta subunit of coupling factor one	2	Col2/C24
At2g39730	auxin-regulated protein	1	Col2/C24
At1g20020	ferredoxin--NADP reductase precursor, putative	3	Col2/C24
At3g08940	putative chlorophyll a/b-binding protein	3	Col2/C24
NP_051055	photosystem II 44 kDa protein	3	Col2/C24
At3g03530	expressed protein, supported by cDNA: gi_14335155	1	Col2/C24
At2g14380	putative retroelement pol polyprotein	3	Col2/C24
At3g26490	Non-phototropic hypocotyl, putative	1	Col2/C24
At5g66520	selenium-binding protein-like	1	Col2/C24
CAB52747	photosystem I subunit III precursor	2	Col2/C24
AAK68813	H+-transporting ATP synthase-like protein	2	Col2/C24
At3g22520	unknown protein	1	Col2/C24
At3g63140	mRNA binding protein precursor - like	1	Col2/C24
At1g79040	photosystem II polypeptide, putative	2	Col2/C24
At5g20290	putative protein	1	Col2/C24
At1g49290	hypothetical protein	1	Col2/C24
At5g13160	protein kinase-like	1	Col2/C24
At5g24630	unknown protein	1	Col2/C24
At1g36990	hypothetical protein	1	Col2/C24
A96754	Similar to part of disease resistance protein [imported]	1	Col2/C24
T05822	hypothetical protein T5K18.170	1	Col2/C24
At3g57330	potential calcium-transporting ATPase 11, plasma membrane-type (Ca2+-ATPase, isoform 11)	1	Col2/C24
At4g25430	hypothetical protein	1	Col2/C24
At1g16240	expressed protein	1	Col2/C24
At1g48280	expressed protein	1	Col2/C24
AAM62447	glycine-rich RNA binding protein 7	1	Col2/C24
BAB08888	gene_id:MIJ24.6~ref|NP_013897.1~similar to unknown protein	1	Col2/C24
At3g14420	glycolate oxidase, putative	1	Col2/C24
AAM65044	60S acidic ribosomal protein P2	1	Col2/C24
At2g29450	glutathione transferase (103-1A)	1	Col2/C24
AAM98072	unknown protein	1	Col2/C24
AF428455_1	putative fructose-bisphosphate aldolase	1	Col2/C24

At3g50820	photosystem II oxygen-evolving complex 33 (OEC33)	1	Col2/C24
At3g14415	glycolate oxidase	1	Col2/C24
At2g20230	expressed protein	1	Col2/C24
At2g13360	alanine-glyoxylate aminotransferase	1	Col2/C24
At5g42650	allene oxide synthase	1	Col2/C24
At2g05100	Light-harvesting chlorophyll a/b binding protein	1	Col2/C24
At5g38750	putative protein	1	Col2/C24
At3g44890	RP19 gene for chloroplast ribosomal protein CL9	1	Col2/C24
At3g45590	putative protein	1	Col2/C24
At5g56810	F-box protein	1	Col2/C24
At1g69070	hypothetical protein	1	Col2/C24
At2g15325	hypothetical protein	1	Col2/C24
At3g63190	putative protein	1	Col2/C24
At1g13800	hypothetical protein	1	Col2/C24
At3g30843	hypothetical protein	1	Col2/C24
S49030	RNA-binding protein RNP-D precursor	1	Col2/C24
AAM66135	unknown	1	Col2/C24
At4g37460	putative protein	1	Col2/C24
At5g44870	disease resistance protein (TIR-NBS-LRR class), putative	1	Col2/C24
At2g47610	60S ribosomal protein L7A	1	Col2/C24
At5g25590	putative protein	1	Col2/C24
AAM63618	putative rubisco subunit binding-protein alpha subunit	1	Col2/C24
At5g23700	putative protein	1	Col2/C24
NP_051045	ATP synthase CF0 B chain	1	Col2/C24
At3g23400	expressed protein	1	Col2/C24
At2g40630	expressed protein	1	Col2/C24
At4g22780	Translation factor EF-1 alpha - like protein	1	Col2/C24
At1g04800	unknown protein	1	Col2/C24
At1g20060	kinesin-related protein	1	Col2/C24
At5g60120	APETALA2 protein - like	1	Col2/C24
At2g01620	expressed protein	1	Col2/C24
At2g27000	cytochrome p450 family	1	Col2/C24
At5g17370	hypothetical protein	1	Col2/C24
At5g09660	microbody NAD-dependent malate dehydrogenase	1	Col2/C24
At4g31050	putative protein	1	Col2/C24
At2g37310	hypothetical protein	1	Col2/C24
At5g49120	putative protein	1	Col2/C24

At3g24550	protein kinase, putative	1	Col2/C24
At2g26940	putative C2H2-type zinc finger protein	1	Col2/C24
T51531	biotin carboxyl carrier protein homolog T20K14.140 [imported]	1	Col2/C24
At5g16860	putative protein	1	Col2/C24
BAB10393	contains similarity to En/Spm-like transposon	1	Col2/C24
At4g18820	putative protein	1	Col2/C24
At5g01730	putative protein	1	Col2/C24
At1g80910	myrosinase precursor, putative	1	Col2/C24
At2g05170	expressed protein	1	Col2/C24
At5g42920	putative protein	1	Col2/C24
At5g51200	putative protein	1	Col2/C24
At3g53720	putative protein	1	Col2/C24
At5g22450	putative protein	1	Col2/C24
At1g22410	3-deoxy-D-arabino-heptulosonate 7-phosphate, putative	1	Col2/C24
At4g30990	putative protein	1	Col2/C24
At3g20860	putative serine/threonine protein kinase	1	Col2/C24
At3g05470	unknown protein	1	Col2/C24
At1g06380	hypothetical protein	1	Col2/C24
At3g47140	putative protein	1	Col2/C24
At5g01630	putative protein	1	Col2/C24
At5g39960	putative protein	1	Col2/C24
At2g35300	similar to late embryogenesis abundant proteins	1	Col2/C24
At4g30830	putative protein	1	Col2/C24
At1g68940	hypothetical protein	1	Col2/C24
At1g79680	WAK-like kinase (WLK)	1	Col2/C24
At3g66658	betaine aldehyde dehydrogenase, putative	1	Col2/C24
At4g19320	hypothetical protein	1	Col2/C24
At3g49350	GTPase activating -like protein	1	Col2/C24
At1g16140	WAK-like kinase (WLK)	1	Col2/C24
At3g04740	hypothetical protein	1	Col2/C24
At3g60890	putative protein	1	Col2/C24
At2g07010	putative retroelement pol polyprotein	1	Col2/C24
At5g02060	putative protein	1	Col2/C24
At3g60310	putative protein	1	Col2/C24
AAA32797	geranylgeranyl pyrophosphate synthase	1	Col2/C24
BAB09274	histidine kinase-like protein	1	Col2/C24
At1g72500	hypothetical protein	1	Col2/C24

At3g42320	putative protein	1	Col2/C24
H86321	hypothetical protein F6A14.10 [imported]	1	Col2/C24
At5g58980	random slug protein - like	1	Col2/C24
At5g14350	putative protein	1	Col2/C24
At4g27720	putative protein	1	Col2/C24
S20866	L-ascorbate peroxidase (EC 1.11.1.11) precursor	3	Col2/C24
At3g62410	CP12 protein precursor-like protein	1	Col2/C24
At3g49120	peroxidase, putative	1	Col2/C24
At3g11540	spindly (gibberellin signal transduction protein)	1	Col2/C24
AAB07026	catalase	3	Col2/C24
At4g33010	P-Protein - like protein	3	Col2/C24
RBS1_ARATH	Ribulose bisphosphate carboxylase small chain 1A, chloroplast precursor (RuBisCO small subunit 1A)	6	Col2/C24
At3g45140	lipoxygenase AtLOX2	7	Col2
S11852	photosystem II oxygen-evolving complex protein 1 precursor	8	Col2
At5g54270	light-harvesting chlorophyll a/b binding protein, putative	3	Col2
JQ1286	glyceraldehyde-3-phosphate dehydrogenase (NADP) (phosphorylating) (EC 1.2.1.13) B precursor, chloroplast	8	Col2
At1g56190	phosphoglycerate kinase, putative	7	Col2
At3g04120	glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	3	Col2
At2g02930	glutathione transferase, putative	5	Col2
AAA50156	carbonic anhydrase	5	Col2
AT4g37930	glycine hydroxymethyltransferase-like protein	5	Col2
At4g04640	coded for by A. thaliana cDNA AA041141	5	Col2
T52072	hypothetical protein g5bf [imported]	5	Col2
NP_051058	photosystem I P700 apoprotein A2	5	Col2
AT3g48870	AtClpC endopeptidase Clp ATP-binding chain C	5	Col2
T12970	hypothetical protein T6H20.190	3	Col2
A96602	elongation factor EF-2 [imported]	5	Col2
At2g39730	Rubisco activase	3	Col2
CAA70862	ferredoxin-dependent glutamate synthase	4	Col2
AF326861_1	putative photosystem I subunit PSI-E	3	Col2
AAN31836	putative 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	4	Col2
AT4g38970	putative fructose-bisphosphate aldolase	2	Col2
T52314	chlorophyll a/b-binding protein Lhcb6 [imported]	4	Col2
At2g35370	glycine decarboxylase complex H-protein	2	Col2

NP_051084	photosystem II 47 kDa protein	2	Col2
At4g13400	putative protein	1	Col2
AAN31832	putative chloroplast translation elongation factor EF-Tu precursor	3	Col2
At5g38410	ribulose bisphosphate carboxylase small chain 3b precursor	3	Col2
AAA32813	plasma membrane proton pump H+ ATPase	3	Col2
NP_051039	photosystem II protein D1	3	Col2
At4g18480	protein ch-42 precursor, chloroplast	3	Col2
S16582	fructose-bisphosphatase (EC 3.1.3.11) precursor, chloroplast	4	Col2
At5g47210	putative protein	2	Col2
JT0901	chaperonin 60 beta precursor	3	Col2
CAA74895	ribosomal protein L4	1	Col2
At1g76030	vacuolar ATP synthase subunit B	2	Col2
S33707	DNA-damage repair protein DRT112 precursor	1	Col2
AAN31859	putative heat shock protein 81-2 (HSP81-2)	3	Col2
AAM63250	cyanate lyase	1	Col2
AF360195_1	putative alanine aminotransferase	2	Col2
At1g53310	phosphoenolpyruvate carboxylase 1, putative	2	Col2
At2g26080	putative glycine dehydrogenase	2	Col2
At1g04410	putative malate dehydrogenase	2	Col2
At1g50250	chloroplast FtsH protease	2	Col2
G84888	probable transketolase precursor [imported]	2	Col2
AAK73957	putative ftsH chloroplast protease	3	Col2
AF083913_1	annexin	2	Col2
At1g77160	hypothetical protein	1	Col2
At3g17820	glutamine synthetase, putative	2	Col2
C84582	hypothetical protein At2g19880 [imported]	1	Col2
At2g41560	potential calcium-transporting ATPase 4, plasma membrane-type (Ca2+-ATPase, isoform 4)	2	Col2
AAM62764	glutamine synthetase, putative	2	Col2
At1g24706	F5A9.21, unknown	1	Col2
At2g22010	unknown protein	1	Col2
At5g14740	carbonic anhydrase 2	1	Col2
At1g76180	dehydrin, putative	2	Col2
AF428455_1	putative fructose-bisphosphate aldolase	1	Col2
At4g31700	ribosomal protein S6 - like	1	Col2
At1g74060	putative 60S ribosomal protein L6	1	Col2
At3g47070	putative protein	2	Col2

At3g58730	v-ATPase subunit D (vATPD)	2	Col2
At1g23740	putative auxin-induced protein	1	Col2
CAA11554	2-oxoglutarate dehydrogenase, E3 subunit	2	Col2
At2g21170	putative triosephosphate isomerase	2	Col2
At3g46520	actin 12	1	Col2
At1g19570	dehydroascorbate reductase, putative	2	Col2
AAM97062	unknown protein	1	Col2
S71112	catalase (EC 1.11.1.6) 3, peroxisome/glyoxysome location signal (S-[RKH]-L) motif	1	Col2
At4g13940	adenosylhomocysteinase	2	Col2
NP_051087	photosystem II phosphoprotein	1	Col2
At2g15620	ferredoxin--nitrite reductase	1	Col2
At3g47470	light-harvesting chlorophyll a/b binding protein	1	Col2
At4g35090	catalase 2	1	Col2
S19226	cold-regulated protein cor47	2	Col2
CAC35872	H+-transporting ATP synthase beta chain (mitochondrial)-like protein	2	Col2
AAB80700	glycolate oxidase	1	Col2
At4g09000	14-3-3 protein GF14 chi (grf1)	1	Col2
At2g34430	photosystem II type I chlorophyll a /b binding protein	1	Col2
At4g34870	peptidylprolyl isomerase (cyclophilin)	2	Col2
At1g16880	expressed protein	2	Col2
At2g37660	expressed protein	2	Col2
AF360228_1	putative glutathione reductase	1	Col2
At3g49910	60S ribosomal protein - like	2	Col2
At5g15980	putative protein	2	Col2
T52122	chaperonin 10	2	Col2
At3g07570	unknown protein	1	Col2
AAM13161	ATP-dependent transmembrane transporter, putative	1	Col2
AAB09585	ADP glucose pyrophosphorylase small subunit	2	Col2
At3g02090	putative mitochondrial processing peptidase	2	Col2
At4g03430	putative pre-mRNA splicing factor	1	Col2
At1g71240	hypothetical protein	1	Col2
At4g31700	ribosomal protein S6 - like	1	Col2
AAK59424	putative DEF (CLA1) protein	1	Col2
At5g55180	glycosyl hydrolase family 17	1	Col2
At1g74770	hypothetical protein	1	Col2

AC012394_17	putative phytochrome A signaling protein	1	Col2
At2g24820	putative Rieske iron-sulfur protein	1	Col2
At2g36380	ABC transporter family protein	1	Col2
At2g35120	glycine decarboxylase complex H-protein	1	Col2
AAL32516	putative protein	1	Col2
At1g50730	hypothetical protein	1	Col2
At2g14470	putative helicase	2	Col2
At1g67560	putative lipoxygenase	1	Col2
At1g56190	phosphoglycerate kinase	1	Col2
At4g12180	putative reverse transcriptase	1	Col2
At3g51560	disease resistance protein (TIR-NBS-LRR class), putative	1	Col2
At1g50120	hypothetical protein	2	Col2
G86301	probable retroelement polyprotein [imported]	1	Col2
At5g16500	protein kinase-like protein	1	Col2
At1g60860	GCN4-complementing protein	1	Col2
CAB80674	putative protein transport factor	1	Col2
At2g34610	hypothetical protein	1	Col2
T01733	hypothetical protein A_IG002N01.31	2	Col2
At2g07698	hypothetical protein	1	Col2
At5g10790	ubiquitin-specific protease 22 (UBP22)	1	Col2
At1g62810	amine oxidase, putative	1	Col2
AC069473_9	unknown protein	1	Col2
At1g73980	unknown protein	1	Col2
T50928	calmodulin-binding protein [imported]	1	Col2
AAM62795	60S ribosomal protein L27A	1	Col2
At5g37670	Low-molecular-weight heat shock protein - like	1	Col2
At5g48010	pentacyclic triterpene synthase (04C11) (ATPEN1), putative	1	Col2
At2g43560	FKBP-type peptidyl-prolyl cis-trans isomerase	1	Col2
AC007354_10	Strong similarity to gb|Y09533 involved in starch metabolism from Solanum tuberosum	1	Col2
At1g13790	hypothetical protein	1	Col2
At2g19380	RRM-containing RNA-binding protein	1	Col2
At5g06240	unknown protein	1	Col2
At1g67240	mutator-like transposase, putative	1	Col2
CAA69802	ATPase subunit 1	1	Col2
At4g20890	tubulin beta-9 chain	1	Col2
At2g40590	40S ribosomal protein S26	1	Col2

BAA97188	emb|CAB87273.1~gene_id:MMI9.7~similar to unknown protein	1	Col2
At5g09860	expressed protein	1	Col2
At1g60630	leucine-rich repeat transmembrane protein kinase	1	Col2
At1g02500	s-adenosylmethionine synthetase	1	Col2
AF462865_1	unknown protein	1	Col2
AF424618_1	membrane-associated salt-inducible protein	1	Col2
At1g05530	UDP-glycosyltransferase family	1	Col2
At1g74680	Exostosin family	1	Col2
At1g32470	glycine cleavage system H protein precursor, putative	1	Col2
At2g47470	putative protein disulfide-isomerase	1	Col2
T48997	epsin-like protein	1	Col2
AAK96795	acyl carrier protein (ACP) gene	1	Col2
At2g16890	putative glucosyltransferase	1	Col2
AAL91646	unknown protein	1	Col2
At5g59660	leucine-rich repeat transmembrane protein kinase, putative	2	Col2
At5g66190	ferredoxin-NADP+ reductase	1	C24
At3g22910	potential calcium-transporting ATPase 13, plasma membrane-type (Ca2+-ATPase, isoform 13)	1	C24
At4g28750	photosystem I subunit PSI-E - like protein	1	C24
At5g48310	putative protein	1	C24
At5g28300	GTL1 - like protein	1	C24
At1g29930	light-harvesting chlorophyll a/b binding protein	1	C24
At5g09660	microbody NAD-dependent malate dehydrogenase	1	C24
At3g11820	syntaxin SYP121	1	C24
At5g40480	nuclear pore protein -like	1	C24
At2g35920	putative ATP-dependent RNA helicase A	1	C24
At1g22490	expressed protein	1	C24
At5g14070	glutaredoxin-like protein	1	C24
NP_051048	ribosomal protein S2	1	C24
At4g19750	glycosyl hydrolase family 18	1	C24
At1g25340	myb-related transcription factor (cpm7), putative	1	C24
At2g25140	HSP100/ClpB, putative	1	C24
At2g20960	pEARLI 4 protein	1	C24
At2g27480	putative calcium binding protein	1	C24
AAD03443	contains similarity to human RNA polymerase II complex component SRB7 (GB:U52960)	1	C24
At5g03940	signal recognition particle 54CP protein precursor	1	C24

At1g07430	protein phosphatase 2C (PP2C), putative	1	C24
At4g07960	putative glucosyltransferase	1	C24
At3g11630	putative 2-cys peroxiredoxin BAS1 precursor (thiol-specific antioxidant protein)	1	C24
At4g01310	putative L5 ribosomal protein	1	C24
At5g61250	glycosyl hydrolase family 79 (endo-beta-glucuronidase/heparanase)	1	C24
At1g65010	hypothetical protein	1	C24
At1g67810	unknown protein	1	C24
At5g50260	cysteine proteinase	1	C24
At5g64040	photosystem I reaction center subunit PSI-N precursor (PSI-N)	1	C24
At5g24770	vegetative storage protein Vsp2	1	C24
At4g28630	ABC transporter family protein	1	C24
At5g09730	glycosyl hydrolase family 3	1	C24
At4g31300	20S proteasome beta subunit A (PBA1);	1	C24
T05498	hypothetical protein T19K4.190	1	C24
AC000103_3	unknown protein	1	C24
At5g09700	beta-glucosidase - like protein	1	C24
At5g15200	40S ribosomal protein - like	1	C24
At1g72300	leucine-rich repeat transmembrane protein kinase, putative	1	C24
At3g14350	leucine-rich repeat transmembrane protein kinase, putative	1	C24
At3g13160	expressed protein	1	C24
At5g59660	leucine-rich repeat transmembrane protein kinase, putative	1	C24
At3g05400	sugar transporter, putative	1	C24
At5g54290	cytochrome c biogenesis protein precursor (gb|AAF35369.1)	1	C24
At1g75350	chloroplast 50S ribosomal protein L31, putative	1	C24
At1g75350	chloroplast 50S ribosomal protein L31, putative	1	C24
At3g54890	light-harvesting chlorophyll a/b binding protein	1	C24
At5g47180	VAMP (vesicle-associated membrane protein)-associated protein-like	1	C24
At5g41610	Na+/H+ antiporter-like protein	1	C24
AC002423_15	unknown protein	1	C24
At5g58490	cinnamoyl-CoA reductase - like protein	1	C24
C86379	unknown protein	1	C24
At1g19640	S-adenosyl-L-methionine:jasmonic acid carboxyl methyltransferase (JMT)	1	C24
At5g04290	glycine-rich protein	1	C24
At1g35680	50S ribosomal protein L21 chloroplast precursor (CL21)	1	C24

NP_051097	ribosomal protein L22	1	C24
At5g48600	chromosome condensation protein	1	C24
At3g43190	sucrose synthase, putative	1	C24
At3g26790	transcriptional regulator (FUSCA3)	1	C24
At4g17300	asparagine--tRNA ligase	1	C24
At1g31000	hypothetical protein	1	C24
BAB02913	unknown protein	1	C24
AAK64154	unknown protein	1	C24
AAB61690	disease resistance protein homolog	1	C24
At2g24490	putative replication protein A1	1	C24
At5g66190	ferredoxin-NADP+ reductase	1	C24
At3g22910	potential calcium-transporting ATPase 13, plasma membrane-type (Ca2+-ATPase, isoform 13)	1	C24
At5g48310	putative protein	1	C24
At5g28300	GTL1 - like protein	1	C24
At1g29930	light-harvesting chlorophyll a/b binding protein	1	C24
At5g09660	microbody NAD-dependent malate dehydrogenase	1	C24
At3g11820	syntaxin SYP121	1	C24
At3g51570	disease resistance protein (TIR-NBS-LRR class), putative	1	C24
At3g46530	disease resistance protein, RPP13-like (CC-NBS class), putative	1	C24
At2g25710	biotin holocarboxylase synthetase	1	C24
At2g12150	Mutator-like transposase	1	C24
At5g32481	Athila retroelement ORF1, putative	1	C24
At3g24190	expressed protein	1	C24
At1g19390	WAK-like kinase (WLK)	1	C24
At4g22470	extensin - like protein	1	C24
At1g47560	hypothetical protein	1	C24
At2g16780	putative WD-40 repeat protein, MSI2	1	C24
At1g18030	protein phosphatase 2C (PP2C), putative	1	C24
At3g23790	AMP-binding protein, putative	1	C24
At1g66530	arginyl-tRNA synthetase	1	C24
At1g48150	MADS-box protein	1	C24
At4g14140	(cytosine-5-)-methyltransferase	1	C24
At1g62940	4-coumarate:coenzyme A ligase, putative	1	C24
At2g29500	putative small heat shock protein	1	C24
At3g07980	putative MAP3K epsilon protein kinase	1	C24
At4g23940	putative MAP3K epsilon protein kinase	1	C24

At1g21810	myosin-like protein	1	C24
At5g14950	glycosyl hydrolase family 38 (alpha-mannosidase)	1	C24
At3g53280	cytochrome P450 monooxygenase	1	C24
At2g41310	putative two-component response regulator 3 protein	1	C24
At1g50410	DNA-binding protein, putative	1	C24
At5g05340	peroxidase, putative	1	C24
A96721	probable peptide transporter	1	C24
At2g26790	putative salt-inducible protein	1	C24
At3g30570	putative reverse transcriptase	1	C24
At1g74080	putative transcription factor	1	C24
At3g21210	CHP-rich zinc finger protein, putative	1	C24
At2g42270	U5 small nuclear ribonucleoprotein helicase, putative	1	C24
At1g69320	CLE10, putative	1	C24
At3g42950	polygalacturonase, putative	1	C24

Table 3: List of the proteins identified in the LC/MS/MS analysis as depicted in Figure 4A and via database search

6.57 e+011	26/112 (23%)	96024.5 / 8.27	>gi|15234315|ref|NP_192920.1| putative phospholipase D-gamma [Arabidopsis thaliana]
3.05 e+011	27/112 (24%)	92351.2 / 6.20	>gi|15240282|ref|NP_200964.1| putative protein [Arabidopsis thaliana]
1.52 e+011	28/112 (25%)	92718.8 / 5.43	>gi|15242062|ref|NP_197578.1| putative protein [Arabidopsis thaliana]
2.24 e+010	24/112 (21%)	86209.1 / 6.18	(AB012239) contains similarity to receptor-like protein kinase~gene_id:K11J9.9
1.82 e+010	22/112 (19%)	73879.2 / 8.66	>gi|4220445|gb|AAD12672.1| (AC006216) Similar to gi|3004555 F19F24.14 salt inducible protein homolog from Arabidopsis thaliana BAC gb|AC003673
4.15 e+009	18/112 (16%)	80138.8 / 5.98	>gi|15221590|ref|NP_176469.1| amine oxidase, putative [Arabidopsis thaliana]
3.42 e+009	19/112 (16%)	44182.4 / 9.25	(AC013258) unknown protein; 14-1201
2.28 e+009	26/112 (23%)	67135.1 / 6.21	>gi|15228343|ref|NP_190398.1| cell division cycle protein 23 homolog [Arabidopsis thaliana]
2.21 e+009	17/112 (15%)	53996.5 / 6.24	>gi|15230447|ref|NP_190700.1| putative protein [Arabidopsis thaliana]
2.21 e+009	24/112 (21%)	79014.5 / 7.84	>gi|15235423|ref|NP_192166.1| putative cullin-like 1 protein [Arabidopsis thaliana]
1.77 e+009	29/112 (25%)	89037.9 / 8.65	>gi|15235498|ref|NP_192184.1| hypothetical protein [Arabidopsis thaliana]
1.5 e+009	16/112 (14%)	89481.9 / 5.78	>gi|15225595|ref|NP_181522.1| putative isoamylase [Arabidopsis thaliana]
1.21 e+009	25/112 (22%)	76115.8 / 9.03	>gi|15219271|ref|NP_175737.1| hypothetical protein [Arabidopsis thaliana]
1.16 e+009	25/112 (22%)	53350.6 / 9.10	(AC001229) Similar to Arabidopsis cytochrome P450 CYP90 (gb|X87367).
1.12 e+009	25/112 (22%)	55128.6 / 8.60	>gi|15218776|ref|NP_176744.1| cytochrome P450, putative [Arabidopsis thaliana]
1.08 e+009	21/112 (18%)	34622.9 / 8.50	>gi|15241862|ref|NP_201058.1| putative protein [Arabidopsis thaliana]
9.98 e+008	23/112 (20%)	70809.8 / 9.19	>gi|15228449|ref|NP_186956.1| putative 26S proteosome regulatory subunit [Arabidopsis thaliana]
9.57 e+008	25/112 (22%)	91655.7 / 9.43	>gi|15242321|ref|NP_196478.1| RNA helicase-like protein [Arabidopsis thaliana]
7.44 e+008	18/112 (16%)	50284.9 / 6.58	>gi|15228443|ref|NP_189791.1| vacuolar H(+)-ATPase subunit-like protein [Arabidopsis thaliana]
7.43 e+008	19/112 (16%)	50357.9 / 6.58	(AY037176) AT3g42050/F4M19_10

7.14 e+008	17/112 (15%)	59440.9 / 9.67	(AC008046) Hypothetical protein
6.34 e+008	17/112 (15%)	94851.5 / 6.35	(AF364174) chromomethylase 3
5.61 e+008	26/112 (23%)	79480.6 / 6.49	>gi|15222507|ref|NP_176554.1| unknown protein [Arabidopsis thaliana]
4.8 e+008	22/112 (19%)	95794.4 / 6.19	(AC007980) Very similar to disease resistance proteins
4.5 e+008	13/112 (11%)	77293.1 / 8.80	>gi|15240114|ref|NP_198530.1| sen1-like protein [Arabidopsis thaliana]
4.47 e+008	21/112 (18%)	95834.2 / 8.32	>gi|15237673|ref|NP_200654.1| putative protein [Arabidopsis thaliana]
4.44 e+008	27/112 (24%)	91155.0 / 8.56	(AJ299417) hypothetical protein
4.44 e+008	13/112 (11%)	78390.5 / 8.70	(AB017068) contains similarity to nonsense-mediated mRNA decay trans-acting factors~gene_id:MJG14.20
3.9 e+008	20/112 (17%)	97275.5 / 6.18	>gi|15222893|ref|NP_175437.1| disease resistance protein, putative [Arabidopsis thaliana]
2.82 e+008	21/112 (18%)	63750.1 / 6.15	(AY056217) putative cell division cycle protein 23
2.57 e+008	24/112 (21%)	73831.6 / 6.43	(AF281655) zeaxanthin epoxidase
2.57 e+008	24/112 (21%)	73842.7 / 6.60	>gi|15240169|ref|NP_201504.1| zeaxanthin epoxidase precursor [Arabidopsis thaliana]
2.56 e+008	29/112 (25%)	92376.5 / 6.53	>gi|15232837|ref|NP_186850.1| hypothetical protein [Arabidopsis thaliana]
2.52 e+008	20/112 (17%)	92074.5 / 6.60	(AF277982) origin recognition complex 1
2.13 e+008	20/112 (17%)	97220.8 / 6.89	(AB052756) SRKb
1.93 e+008	22/112 (19%)	73984.2 / 7.50	>gi|15221081|ref|NP_172636.1| putative salt-inducible protein [Arabidopsis thaliana]
1.91 e+008	20/112 (17%)	90361.3 / 6.01	>gi|15225790|ref|NP_180247.1| putative salt-inducible protein [Arabidopsis thaliana]
1.79 e+008	26/112 (23%)	73268.2 / 5.11	>gi|15229907|ref|NP_187164.1| hypothetical protein [Arabidopsis thaliana]
1.77 e+008	15/112 (13%)	80208.1 / 5.72	>gi|15241843|ref|NP_198205.1| far-red impaired response protein (FAR1) - like [Arabidopsis thaliana]
1.73 e+008	19/112 (16%)	81148.9 / 8.74	(AP000377) long-chain-fatty-acid CoA ligase
1.73 e+008	25/112 (22%)	79706.0 / 7.19	>gi|15223594|ref|NP_176062.1| hypothetical protein [Arabidopsis thaliana]
1.7 e+008	18/112 (16%)	63806.0 / 5.93	>gi|15229518|ref|NP_189021.1| AMP-binding protein, putative [Arabidopsis thaliana]
1.63 e+008	20/112 (17%)	53553.2 / 11.48	>gi|15225620|ref|NP_181536.1| En/Spm-like transposon protein [Arabidopsis thaliana]

1.59 e+008	23/112 (20%)	98473.6 / 5.64	>gi|15239960|ref|NP_199187.1| disease resistance protein [Arabidopsis thaliana]
1.5 e+008	24/112 (21%)	73854.7 / 6.60	(AB030296) AtABA1
1.47 e+008	21/112 (18%)	58491.7 / 5.71	>gi|15242019|ref|NP_200513.1| lycopene epsilon cyclase [Arabidopsis thaliana]
1.47 e+008	21/112 (18%)	58515.7 / 5.79	LYCOPENE EPSILON CYCLASE, CHLOROPLAST PRECURSOR
1.2 e+008	20/112 (17%)	88237.6 / 9.02	>gi|15220762|ref|NP_176422.1| hypothetical protein [Arabidopsis thaliana]
1.18 e+008	20/112 (17%)	89309.9 / 9.07	(AC003113) F24O1.4
1.15 e+008	22/112 (19%)	82749.0 / 5.21	>gi|15240560|ref|NP_199794.1| putative protein [Arabidopsis thaliana]
1.08 e+008	15/112 (13%)	69816.5 / 6.26	>gi|15227604|ref|NP_180530.1| anthranilate synthase, alpha subunit [Arabidopsis thaliana]
1.08 e+008	15/112 (13%)	69880.6 / 6.34	anthranilate synthase (EC 4.1.3.27) alpha-2 chain
1.04 e+008	14/112 (12%)	18350.4 / 4.18	>gi|15221030|ref|NP_173259.1| calcium-binding protein, putative [Arabidopsis thaliana]
1.01 e+008	17/112 (15%)	69901.5 / 4.97	>gi|15218356|ref|NP_175017.1| hypothetical protein [Arabidopsis thaliana]
9.4 e+007	27/112 (24%)	83768.8 / 6.06	>gi|15227316|ref|NP_179280.1| putative salt-inducible protein [Arabidopsis thaliana]
9.34 e+007	15/112 (13%)	94906.6 / 6.37	(AF383170) chromomethylase CMT3
9.24 e+007	19/112 (16%)	86999.6 / 9.41	hypothetical protein F24O1.3
9,00 E+07	14/112 (12%)	62891.3 / 9.76	>gi|15242752|ref|NP_201144.1| auxin-independent growth promoter-like protein [Arabidopsis thaliana]
8.77 e+007	18/112 (16%)	87505.3 / 6.84	>gi|15234605|ref|NP_194731.1| awaiting functional assignment [Arabidopsis thaliana]
8.29 e+007	14/112 (12%)	77819.7 / 6.00	delta1-pyrroline-5-carboxylate synthetase [imported]
8.07 e+007	26/112 (23%)	64010.5 / 6.57	>gi|15228825|ref|NP_188906.1| hypothetical protein [Arabidopsis thaliana]
8.05 e+007	22/112 (19%)	57561.3 / 5.97	glutathione synthase (EC 6.3.2.3) 2 precursor, chloroplast
7.93 e+007	21/112 (18%)	96451.3 / 9.06	(AP000417) beta-1,4-xylosidase
7.41 e+007	17/112 (15%)	45861.3 / 5.51	>gi|15219628|ref|NP_176807.1| unknown protein [Arabidopsis thaliana]
7.1 e+007	14/112 (12%)	94905.6 / 6.57	>gi|15222449|ref|NP_177135.1| putative chromomethylase [Arabidopsis thaliana]
6.97 e+007	13/112 (11%)	41782.4 / 5.12	(AF418279) 2-oxoglutarate-dependent dioxygenase

6.96 e+007	13/112 (11%)	41892.5 / 5.19	(AF418283) 2-oxoglutarate-dependent dioxygenase
6.93 e+007	16/112 (14%)	42066.7 / 5.25	(AF418277) 2-oxoglutarate-dependent dioxygenase
6.91 e+007	13/112 (11%)	42193.8 / 5.19	(AF418278) 2-oxoglutarate-dependent dioxygenase
6.65 e+007	21/112 (18%)	68357.3 / 5.22	>gi|15223533|ref|NP_176036.1| heat shock protein, putative [Arabidopsis thaliana]
6.62 e+007	16/112 (14%)	92104.5 / 6.67	>gi|15219046|ref|NP_175671.1| hypothetical protein [Arabidopsis thaliana]
6.55 e+007	22/112 (19%)	85251.5 / 6.12	>gi|15220477|ref|NP_176925.1| putative protein kinase [Arabidopsis thaliana]
6.49 e+007	16/112 (14%)	44914.1 / 5.26	(AF418276) 2-oxoglutarate-dependent dioxygenase
6.48 e+007	16/112 (14%)	44936.1 / 5.40	(AF418274) 2-oxoglutarate-dependent dioxygenase
6.48 e+007	16/112 (14%)	44943.1 / 5.26	(AF417859) AOP3
6.48 e+007	13/112 (11%)	44944.1 / 5.20	(AF418275) 2-oxoglutarate-dependent dioxygenase
6.12 e+007	26/112 (23%)	97742.7 / 8.10	>gi|15236576|ref|NP_192612.1| putative protein [Arabidopsis thaliana]
5.96 e+007	21/112 (18%)	59364.9 / 9.45	>gi|15236834|ref|NP_194398.1| putative protein [Arabidopsis thaliana]
5.83 e+007	14/112 (12%)	49681.4 / 8.54	>gi|15241740|ref|NP_198757.1| unknown protein [Arabidopsis thaliana]
5.79 e+007	22/112 (19%)	73789.6 / 6.43	(AF134577) zeaxanthin epoxidase
5.22 e+007	26/112 (23%)	99549.4 / 8.45	(AC011708) putative DNA gyrase A subunit
5.2 e+007	13/112 (11%)	74888.6 / 6.92	>gi|15238566|ref|NP_200801.1| putative protein [Arabidopsis thaliana]
5.04 e+007	21/112 (18%)	74663.7 / 5.76	>gi|15234413|ref|NP_194549.1| putative protein [Arabidopsis thaliana]
5.01 e+007	18/112 (16%)	81414.5 / 6.08	>gi|15236567|ref|NP_192606.1| putative MuDR-like transposon protein [Arabidopsis thaliana]
4.87 e+007	16/112 (14%)	81078.9 / 5.65	>gi|15234017|ref|NP_194215.1| brefeldin A-sensitive Golgi protein - like [Arabidopsis thaliana]
4.75 e+007	20/112 (17%)	92699.6 / 8.31	(AF138281) phospholipase D-gamma-2
4.66 e+007	16/112 (14%)	84654.0 / 5.53	(AF367300) AT4g24840/F6I7_50
4.63 e+007	19/112 (16%)	65466.2 / 8.37	(AF369565) membrane protein Mlo4
4.61 e+007	14/112 (12%)	79492.1 / 8.17	>gi|15238268|ref|NP_196088.1| putative protein [Arabidopsis thaliana]

4.53 e+007	22/112 (19%)	79498.0 / 7.20	helicase homolog T6H20.10
4.47 e+007	16/112 (14%)	89843.5 / 7.57	>gi|15241340|ref|NP_196925.1| receptor protein kinase-like protein [Arabidopsis thaliana]
4.33 e+007	19/112 (16%)	78944.3 / 9.20	>gi|15230744|ref|NP_189651.1| hypothetical protein [Arabidopsis thaliana]
4,00 E+07	16/112 (14%)	62984.6 / 8.87	>gi|15232344|ref|NP_188709.1| unknown protein [Arabidopsis thaliana]
3.94 e+007	14/112 (12%)	38665.0 / 4.84	>gi|15240470|ref|NP_200326.1| putative protein [Arabidopsis thaliana]
3.84 e+007	17/112 (15%)	79023.3 / 6.04	(AC069251) F2D10.23

References

ap Rees, T. and Hill S.A. 1994. Metabolic control analysis of plant metabolism. *Plant Cell Env.* **17**: 587-599

Arkin et al. (1997). A test case of correlation metric construction of a reaction pathway from measurements. Science 277, 1275-1279

Benthin, B., Danz, H. and Hamburger, M. 1999. Pressurized liquid extraction of medicinal plants. *J. Chromatogr. A* **837**: 211-219

Blanch, G.P., Caja, M.M., del Castillo, M.L.R., Santa-Mariá, G. and Herraiz, M. 1999. Fractionation of plant extracts by supercritical fluid extraction and direct introduction in capillary gas chromatography using a programmable temperature vaporizer. *J. Chromatogr. Sci.* **37**: 407-410

Bondarenko et al. (2002). Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tendem mass spectrometry. Analytical Chemistry 74, 4741-4749

Bouchereau A., Guénot P. and Larher, F. 2000. Analysis of amines in plant materials. *J. Chromatogr. B* **747**: 49-67

Boyes et al. (2001). Growth stage-based phenotypic analysis of Arabidopsis: A model for high throughput functional genomics in plants. Plant Cell 13, 1499-1510

Castioni, P., Christen, P. and Veuthey J.L. 1995. Supercritical fluid extraction of compounds from plant origin. *Analusis* **23**: 95-106

Chelius et al. (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. Journal of Proteome Research 1, 317-323

Conrads TP, Anderson GA, Veenstra TD, Pasa-Tolic L, Smith RD (2000) Utility of accurate mass tags for proteome-wide protein identification. Anal. Chem. 72, 3349-3354

Eggeling F V, Roemer, I., Dahse, R., Riese, U., Fiedler, W., Ernst, G., Claussen, U. and Klose, J. (1998) Multicolor two-dimensional gel electrophoresis for mutation detection in cancer related genes. J. Molecular Medicine 76 B48.

Farré, E.M., Tiessen, A., Roessner, U., Geigenberger, P., Trethewey, R.N. and Willmitzer, L. 2001. Analysis of the compartmentation of glycolytic intermediates, nucleotides, sugars, organic acids, amino acids and sugar alcohols in potato tubers using a non-aqueous fractionation method. *Plant Physiol. 127, 685-700*

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. 2000. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**: 1157-1161

Fiehn O. (2001) Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol 48, 155-171*

Gámiz-Gracia, L. and de Castro, M.D.L. 2000. Continuous subcritical water extraction of medicinal plant essential oil: comparison with conventional techniques. *Talanta* **51**: 1179-1185

Gassen, H. G. and Schrimpf, G. (1999) Gentechnische Methoden, Spektrum Akademisher Verlag, Gustav Fischer, Heidelberg Berlin

Gerhardt, R. and Heldt, H.W. 1984. Measurement of subcellular metabolite levels in leaves by fractionation of freeze-stopped material in nonaqueous media. *Plant Physiol.* **75**: 542-547

Gonzalez, B., Francois, J. and Renaud, M. 1997. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* **13**: 1347-1356

Goodlett DR, Keller A, Watts JD, Newitt R, Yi EC, Purvine S, Eng JK, von Haller P, Aebersold R, Kolker E (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid. Commun. Mass Spectrom. 15, 1214-1221

Goshe MB, Conrads TP, Panisko EA, Angell NH, Veenstra TD, Smith RD (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. Anal. Chem. 73, 2578-2586

Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology PNAS 97, 9390-9395

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999b) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol.17, 994-999

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999a) Correlation between protein and mRNA abundance in yeast. Mol. Cell Biol. 19, 1720-1730

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai HY, He YDD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH 2000 Functional discovery via a compendium of expression profiles, Cell 102, 109-126

Jarvis, A.P. and Morgan, E.D. 1997. Isolation of plant products by supercritical fluid extraction. *Phytochem. Anal.* **8**: 217-222

Jensen PK, Pasa-Tolic L, Peden KK, Martinovic S, Lipton MS, Anderson GA, Tolic N, Wong KK, Smith RD (2000) Mass spectrometic detection for capillary isoelectric focusing separations of complex protein mixtures. Electrophoresis 21, 1372-1380

Johansen, H.N., Glitso, V. and Knudsen, K.E.B. 1996. Influence of extraction solvent and temperature on the quantitative determination of oligosaccharides from plant materials by high performance liquid chromatography. *J. Agric. Food Chem.* **44**: 1470-1474

Kell et al. (2000). Snapshots of systems: Metabolic control analysis and biotechnology in the postgenomic era. In: Technical and medical implications of metabolic control analysis, ed. AJ Cornish-Bowden, ML Cardenas, pp. 2-25 Netherland: Kluwer Academic Publishers

Koller et al. (2002). Proteomic survey of metabolic pathways in rice. Proc. Natl. Acad. Sci. U.S.A. 99, 11969-11974

Krapp et al. (1993). Regulation of the expression of Rbcs and other photosynthetic genes by carbohydrates – a mechanism for the sink regulation of photosynthesis. Plant J. 3, 817-828

Li LJ, Masselon CD, Anderson GA, Pasa-Tolic L, Lee SW, Shen YF, Zhao R, Lipton MS, Conrads TP, Tolic N, Smith RD (2001) High-throughput peptide identification from protein digests using data-dependent multiplexed tandem FTICR mass spectrometry coupled with capillary liquid chromatography. Anal. Chem. 73, 3312-3322

Link, A, J. 1999. 2D Proteome Analysis. In: Methods in Molecular Biology, Volume 112, Humana Press, Totowa, New Jersey

Lockhart DJ, Winzeler EA Genomics, gene expression and DNA arrays 2000 Nature 405, 827-836.

Namiesnik, J. and Gorecki, T. 2000. Sample preparation for chromatographic analysis of plant material. *J. Plan. Chromatogr.* **13**: 404-413

Nock, C., Gauss, C., Himmelbauer, H., Buessow, K., Schalkwyk, L., Loewe, M., Lehrach, H. and Klose J (1998) Connecting the mouse proteome to its genome. J. Molecular Medicine 76 B46.

Pang et al. (2002). Biomarker discovery in urine by proteomics. J. Proteome Res. 1, 161-169

Ravasz et al. (2002). Hierarchical organization of modularity in metabolic networks. Science 297, 1551-1555

Roberts, J.K.M. 2000. NMR adventures in the metabolic labyrinth within plants. Trends Plant Sci. 5: 30-34

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A.R. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13: 11-29

Sambrook and Russell (2001) Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press, New York

Sargenti, S.R. and Vichnewski, W. 2000. Sonication and liquid chromatography as a rapid technique for extraction and fractionation of plant material. *Phytochem. Anal.* 11: 69-73

Shen YF, Tolic N, Zhao R, Pasa-Tolic L, Li LJ, Berger SJ, Harkewicz R, Anderson GA, Belov ME, Smith RD (2001a) High-throughput proteomics using high efficiency multiple-capillary liquid chromatography with on-line high-performance ESI FTICR mass spectrometry. Anal. Chem. 73, 3011-3021

Shen YF, Zhao R, Belov ME, Conrads TP, Anderson GA, Tang KQ, Pasa-Tolic L, Veenstra TD, Lipton MS, Udseth HR, Smith RD (2001b) Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. Anal. Chem. 73, 1766-1775

Smith RD, Pasa-Tolic L, Lipton MS, Jensen PK, Anderson GA, Shen YF, Conrads TP, Udseth HR, Harkewicz R, Belov ME, Masselon C, Veenstra TD (2001) Rapid quantitative

measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry. Electrophoresis 22, 1652-1668

Starmans, D.A.J. and Nijhus H.H. 1996. Extraction of secondary metabolites from plant material: a review. *Trends Food Sci. Technol.* **7**: 191-197

Streeter J.G. and Strimbu C.E. 1998. Simultaneous extraction and derivatization of carbohydrates from green plant tissues for analysis by gas-liquid chromatography. *Anal. Biochem.* **259**: 253-257

Tweeddale, H., Notley-McRobb, L. and Ferenci, T. 1998. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("Metabolome") analysis. *J. Bacteriol.* **180**: 5109-5116

Wall DB, Kachman MT, Gong SY, Hinderer R, Parus S, Misek DE, Hanash SM, Lubman DM (2000) Isoelectric focusing nonporous RP HPLC: A two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. Anal. Chem. 72, 1099-1111

Wall DB, Kachmann MT, Gong SY, Parus SJ, Long, MW, Lubman DM (2001) Isoelectric focusing nonporous silica RP HPLC/ESI TOF MS: a three dimensional liquid-phase protein separation method as applied to human erythroleukemia cell-line. Rapid Commun. Mass Spectrom. 18, 1649-1661

Washburn MP, Wolters D, Yates JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242-247

Weckwerth et al. (2001). Metabolomic characterization of transgenic potato plants using GC/TOF and LC/MS analysis reveals silent metabolic phenotypes. Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, 1-2

Wolters et al. (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683-5690

Claims

1. A method for providing data useful for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material, whereby said analysis involves suitable statistical evaluation and correlation analysis on the data obtained,
said method being characterized by the step of extracting, identifying and quantifying at least two compound classes of the group consisting of metabolites, proteins and RNA from at least one sample from said biological source material, wherein the compounds of said at least two classes are each determined from one sample.

2. The method of claim 1, wherein said analysis furthermore involves network analysis.

3. The method of claim 1 or 2, wherein extracting, identifying and quantifying is carried out taking a multitude of samples.

4. The method of any on of claims 1 to 3, wherein said extracting comprises the steps of:
   (a) extracting the metabolites from said sample with at least one solvent or mixture of solvents;
   (b) extracting the proteins from the remainder of the sample after step (a);
   (c) extracting the RNA from the remainder of the sample after step (a); and
   (d) optionally dissolving remaining cellular material contained in said sample.

5. The method of claim 4, wherein said mixture of solvents comprises at least one highly polar solvent, at least one less polar solvent and at least one lipophilic solvent.

6. The method of claim 5, wherein said mixture of solvents comprises water, methanol and chloroform.

7. The method of claim 6, wherein said mixture of solvents contains water, methanol and chloroform in the approximate proportion by volume of 1: 2.5: 1.

8. The method of any one of claims 4 to 7, wherein step (a) is carried out at a temperature between -60 and +4°C.

9. The method of any one of claims 1 to 8, further comprising removing detection-disturbing compounds from the metabolites, the polypeptides and/or the RNA prior to identifying and quantifying the metabolites, proteins and/or RNA.

10. The method of claim 9, wherein the detection-disturbing compounds are carbohydrates or other compounds that disturb identification and quantification of RNA.

11. A method for quantitatively analyzing metabolites, proteins and/or RNA in a biological source material comprising
    (a) providing data on metabolites, proteins and/or RNA in said biological source material according to the method of any one of claims 1 to 10;
    (b) performing suitable statistical evaluation and correlation analysis on the data obtained; and
    (c) optionally further performing a network analysis on the data obtained in step (b).

12. Use of a mixture of solvents as defined in any one of claims 4 to 8 for extracting metabolites from a sample of a biological source material in order to perform metabolite profiling.

13. The use of claim 12, wherein additionally proteins and/or RNA is extracted from said sample.

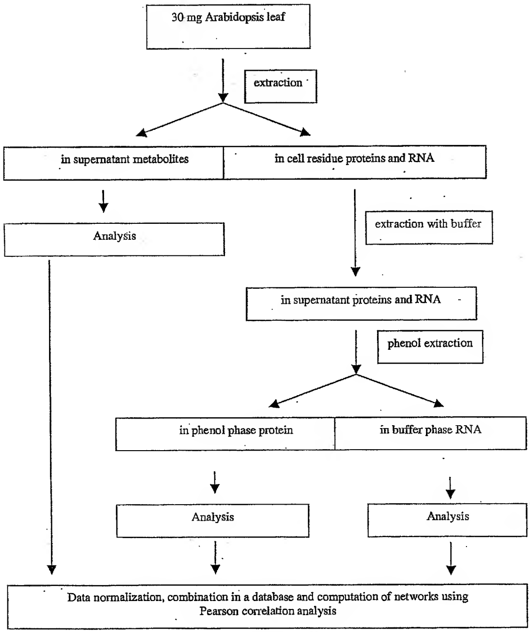
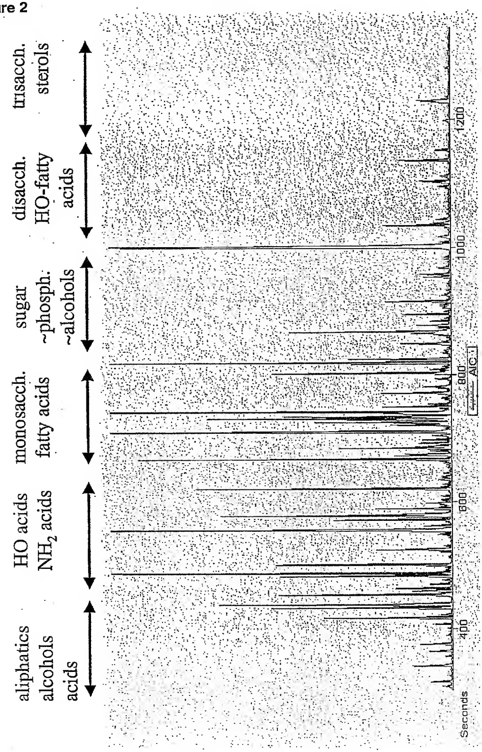

30 mg Arabidopsis leaf
extraction
in supernatant metabolites
in cell residue proteins and RNA
Analysis
extraction with buffer
in supernatant proteins and RNA
phenol extraction
in phenol phase protein
in buffer phase RNA
Analysis
Analysis
Data normalization, combination in a database and computation of networks using Pearson correlation analysis

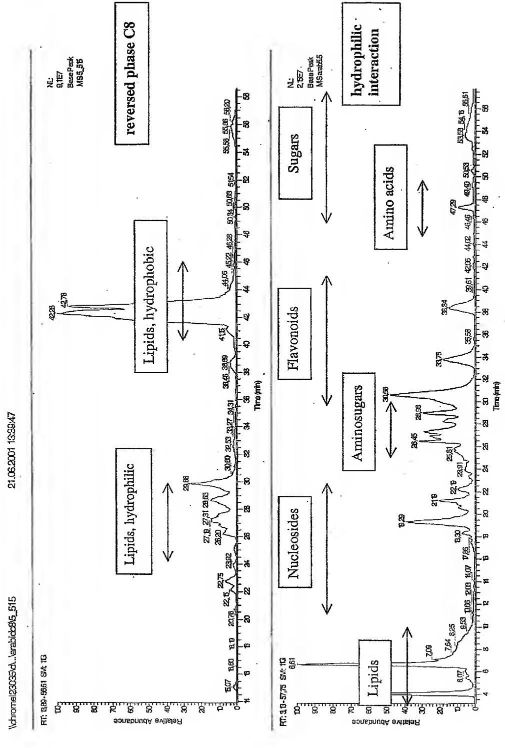
Figure 1

Figure 2


aliphatics
alcohols
acids
HO acids
NH2 acids
monosacch.
fatty acids
sugar
~phosph.
~alcohols
disacch.
HO-fatty
acids
trisacch.
sterols
Seconds
400
600
800
1000
1200
AIC

A

Figure 2 (continued)



B

Figure 3

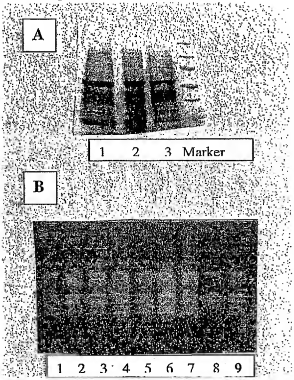


Figure 4 A

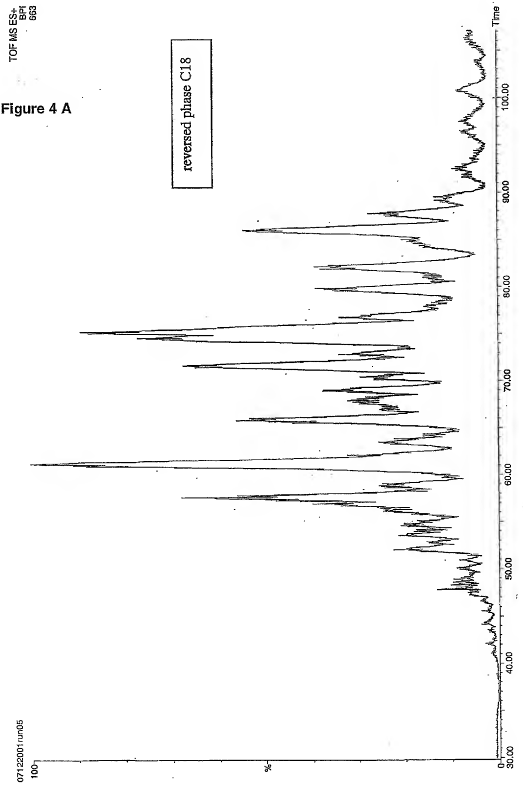

TOF MS ES+
BPI
663
reversed phase C18
07122001run05
100
%
0
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
Time

Figure 4 (continued)

B

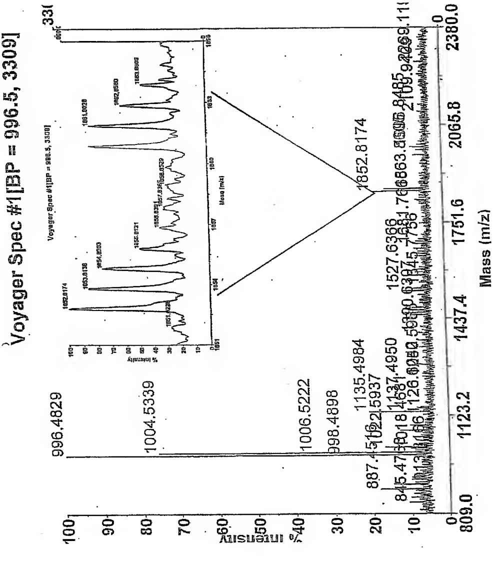
Mass D0 (control)	Mass D8/16 (sample)	Area D0	Area D8/16	ICATRatio
3323.1011	3339.1708	1303.3687	1356.7689	1.04
2501.8528	2509.8861	1978.8151	1590.2736	0.80
2372.7542	2380.7783	1244.9980	1073.7796	0.86
2157.8434	2165.8513	1553.2908	2095.0156	1.35
1985.7446	1993.7988	1642.7697	1381.7960	0.84
1852.6848	1860.7579	2892.8428	4115.1687	1.42
1540.6463	1548.6775	2384.4290	1586.3890	0.67
1464.6553	1472.7028	1369.9453	4197.7218	3.06
1456.5886	1472.7028	2162.8477	4397.6134	2.03
1456.5886	1464.6553	2162.8477	1438.4426	0.67
1251.5166	1259.5456	1312.4298	1644.5472	1.25
1207.5666	1215.6580	2851.2727	1619.7582	0.57
1148.4514	1156.4795	1613.5940	1716.8097	1.06
1108.4930	1116.5416	2195.3149	2335.6855	1.06
1076.4144	1092.5413	3571.4480	2771.2577	0.78
887.4064	903.4846	6653.9668	4653.3889	0.70
833.4152	841.4347	5442.1245	3641.3108	0.67

Average	Std.Dev.
1.07	0.63

WO 03/058238 PCT/EP03/00196

Figure 4 (continued)

C



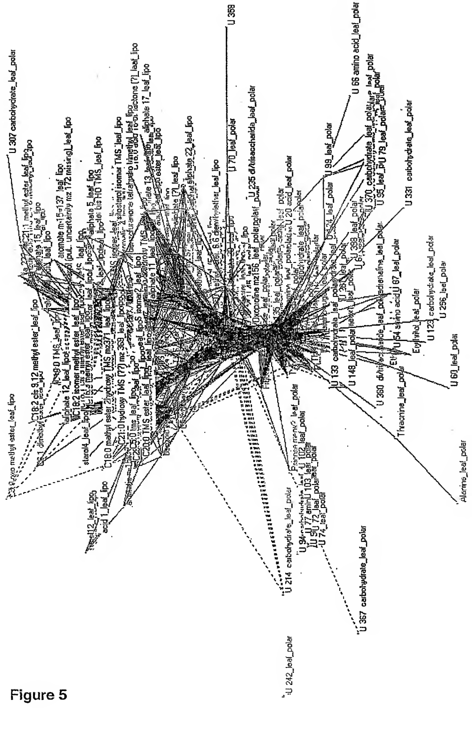
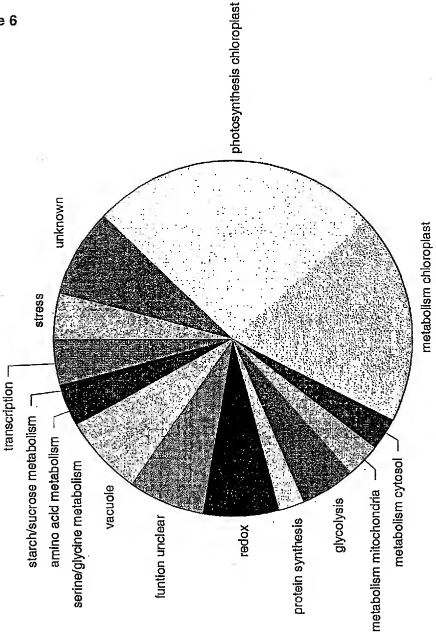

U 307 carbohydrate_leaf_polar
U 66 amino acid_leaf_polar
U 70_leaf_polar
U 235 di/trisaccharide_leaf_polar
U 99_leaf_polar
U 331 carbohydrate_leaf_polar
U 256_leaf_polar
U 60_leaf_polar
Threonine_leaf_polar
Alanine_leaf_polar
U 214 carbohydrate_leaf_polar
U 367 carbohydrate_leaf_polar
U 242_leaf_polar

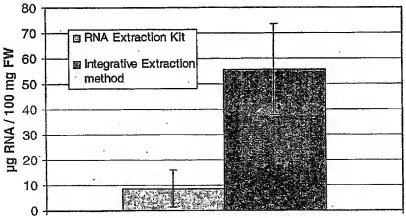
Figure 5

Figure 6

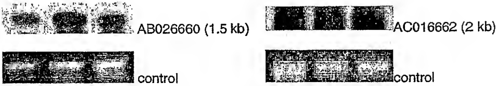

photosynthesis chloroplast
metabolism chloroplast
metabolism cytosol
metabolism mitochondria
glycolysis
protein synthesis
redox
funtion unclear
vacuole
serine/glycine metabolism
amino acid metabolism
starch/sucrose metabolism
transcription
stress
unknown

A

Figure 6 (continued)

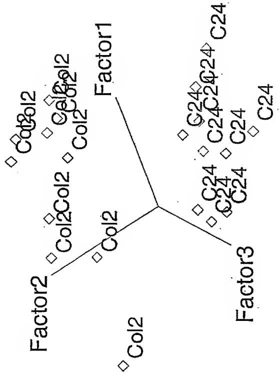


B



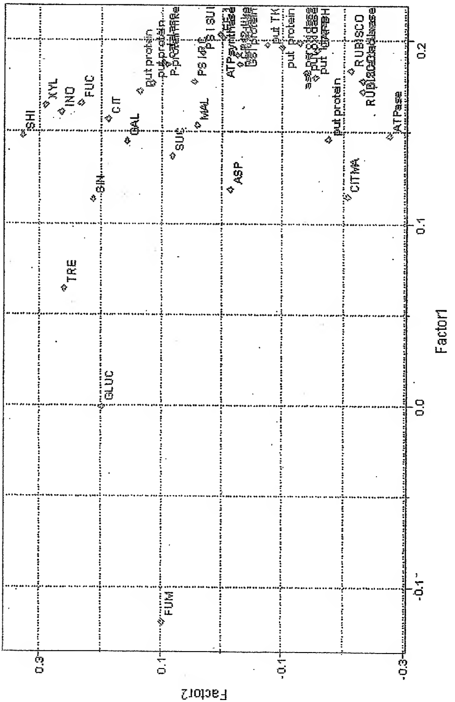
C

Figure 7



A

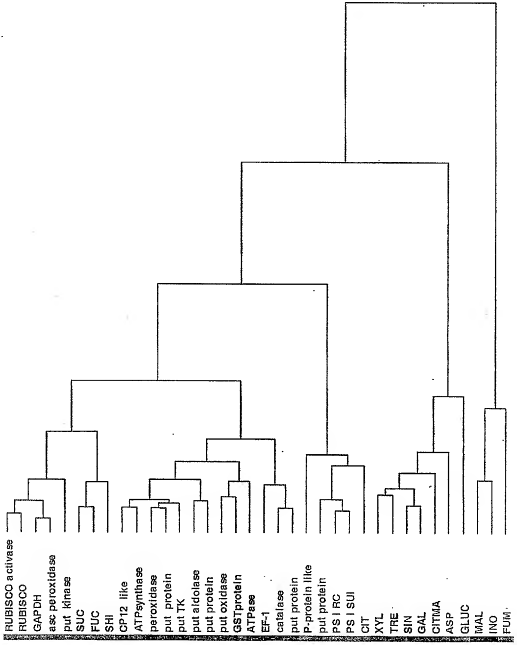
Figure 7 (continued)



B

Figure 7 (continued)

C



INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 03/00196

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G01N33/50 G01N33/68 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, MEDLINE, EMBASE, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FIEHN O ET AL: "COMIBNING GENOMICS, METABOLOME ANALYSIS, AND BIOCHEMICAL MODELLING TO UNDERSTAND METABOLIC NETWORKS" COMPARATIVE AND FUNCTIONAL GENOMICS, XX, XX, vol. 2, no. 3, June 2001 (2001-06), pages 155-168, XP002238133 ISSN: 1531-6912 cited in the application page 157, column 2, paragraph 2 -page 158, column 1, paragraph 2 ---	1-13

-/--

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 June 2003	23/06/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Schalich, J

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CELIS A J E ET AL: "Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics" FEBS LETTERS, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 480, no. 1, 25 August 2000 (2000-08-25), pages 2-16, XP004337487 ISSN: 0014-5793	1,3,4
Y	page 14, paragraph 4 page 4, paragraph 2.5	1-13
X	ORNTOFT TORBEN F ET AL: "Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas." MOLECULAR & CELLULAR PROTEOMICS, vol. 1, no. 1, January 2002 (2002-01), pages 37-45, XP008015037 January, 2002 ISSN: 1535-9476 page 37 -page 45; figures 2,4	1,3,4,11
X	PAULSON L ET AL: "The combination of comperative brain proteome- and cDNA micro array analysis in MK-801 treated rats: An animal model for schizophrenia." JOURNAL OF NEUROCHEMISTRY, vol. 78, no. Supplement 1, September 2001 (2001-09), page 79 XP008015165 Eighteenth Biennial Meeting of the International Society for Neurochemistry and the Thirty-Second Annual Meeting of the American Society for Neurochemistry;Buenos Aires, Argentina; August 26-31, 2001 ISSN: 0022-3042 abstract	1,3,4
P,X	PAULSON LINDA ET AL: "Comparative genome- and proteome analysis of cerebral cortex from MK-801-treated rats." JOURNAL OF NEUROSCIENCE RESEARCH. UNITED STATES 15 FEB 2003, vol. 71, no. 4, 15 February 2003 (2003-02-15), pages 526-533, XP008015233 ISSN: 0360-4012 the whole document	1,3,4

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TYAGI RAKESH KUMAR ET AL: "Stimulation of fructose 1,6-bisphosphate production in melanoma cells by alpha-melanocyte-stimulating hormone 31P/13C-NMR and 32P-labeling studies." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 258, no. 1, November 1998 (1998-11), pages 68-77, XP002235553 ISSN: 0014-2956	12
Y	page 69, column 2 -page 70, column 1	1-13
X	HAEUSLER RAINER E ET AL: "Determination of low-abundant metabolites in plant extracts by NAD(P)H fluorescence with a microtiter plate reader." ANALYTICAL BIOCHEMISTRY, vol. 281, no. 1, 15 May 2000 (2000-05-15), pages 1-8, XP002235554 ISSN: 0003-2697 page 2, column 1, paragraph 4 -page 3, column 1, paragraph 2	7,12
A	FIEHN OLIVER ET AL: "Integrated studies on plant biology using multiparallel techniques." CURRENT OPINION IN BIOTECHNOLOGY, vol. 12, no. 1, February 2001 (2001-02), pages 82-86, XP008014995 ISSN: 0958-1669 cited in the application the whole document	1-13
X	GROPPE JAY C ET AL: "Isolation of full-length RNA templates for reverse transcription from tissues rich in RNase and proteoglycans." ANALYTICAL BIOCHEMISTRY, vol. 210, no. 2, 1993, pages 337-343, XP002243002 ISSN: 0003-2697 page 338, column 1 page 339, column 1, paragraph 2	9,10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 03/00196

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 11
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(i) PCT - Mathematical method

2. [X] Claims Nos.: -
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

International Application No. PCT/EP 03 /00196

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Present claims 9 and 10 relate to a method defined by reference to a desirable characteristic or property, namely removing detection-disturbing compounds from the metabolites, the polypeptides and/or (especially) RNA prior to identification and quantification. The claims cover all methods having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only one such method. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the method by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely methods comprising the removal of carbohydrates, that disturb identification and quantification of RNA .

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.